Digitized by the Internet Archive
in 2024

AUSTRALIAN JOURNAL

OF PHYSICS

VOLUME 6

CONTENTS

NUMBER 1, MARCH 1953

iv

CONTENTS

Number 2, June 1953

PAGE

Conduction of Heat in the Semi-infinite Solid, with a Short Table of an
Important Integral. By R. C. T. Smith 127

On the Standard Errors in the Fitting of Polynomials to Unequally
Spaced Observations. By P. G. Guest 131

Calculation of Accuracy of Results of Graphical Square Intercomparisons.
By P. M. Gilet and G. S. Watson 155

Energy Resolution of γ-Radiation up to 18 MeV by Sodium Iodide
Scintillation Counters. By J. G. Campbell and A. J. F. Boyle .. 171

The Distribution of Radio Brightness over the Solar Disk at a Wavelength
of 21 Centimetres. I. A New Highly Directional Aerial System.
By W. N. Christiansen and J. A. Warburton 190

Further Studies of Electrode Phenomena in Transient Arcs. By W. R.
Blevin 203

Climatic Change in Australia since 1880. By E. L. Deacon 209
The Southern Auroral Zone as Defined by the Position of Homogeneous
Arcs. By F. Jacka 219

Radar Observations of Rain at Sydney, New South Wales. By G. A.
Day 229

NUMBER 3, SEPTEMBER 1953

Short Communications

NUMBER 4, DECEMBER 1953

Short Communications

THE RELATIVISTIC ELECTROMAGNETIC EQUATIONS IN A MATERIAL MEDIUM

By N. W. TAYLOR*

Summary

It is shown how the general relativistic electromagnetic equations for a material medium can be expressed in the form of a single four-vector density equation. The field tensor has six different complex components instead of three, as in the case of a free medium. The classical equations are obtained by separating the real and imaginary parts.

I. INTRODUCTION

It has already been shown how, for empty space, Maxwell's equations in General Relativity may be expressed as a single complex vector density equation from which the complete set of classical equations can be deduced by equating the real and imaginary parts (Taylor 1952). It was pointed out (loc. cit., Section VI) that this seems to be the natural generalization of the quaternionic form used by Silberstein (1924, pp. 46, 206) for classical and special relativistic theory. The purpose of the present investigation is to extend this method of the complex field components to the case of a material medium, for which

$$\mathbf{D} = \varepsilon\mathbf{E}, \quad \mathbf{B} = \mu\mathbf{H}.$$

Some work related to this topic has been attempted before. Following the success of the quaternionic notation in dealing with the electromagnetic equations for a vacuum in classical theory and in Special Relativity, Silberstein (1907) tried a similar treatment for material media. He considered the classical equation (our notation)

$$\frac{\partial\eta}{\partial t} = -iv \operatorname{curl} \eta,$$

where

$$\eta = \sqrt{\varepsilon}\mathbf{E} + i\sqrt{\mu}\mathbf{H}$$

and

$$v = c/\sqrt{\varepsilon\mu}.$$

This is obviously unsuitable for inclusion in a relativistic theory, and besides, the equations div $\mathbf{B} = 0$, div $\mathbf{D} = \rho$ cannot be brought into this scheme. Using a different line of attack (1924, p. 260) he obtained the complete set of Maxwell's equations for a ponderable medium in Special Relativity in terms of *two* quaternionic equations involving the two different bivectors (our notation)

$$\mathbf{B} \pm i\mathbf{E}, \quad \mathbf{H} \pm i\mathbf{D}. \quad \dots\dots\dots\dots\dots\dots\dots \text{(a)}$$

* Department of Mathematics, New England University College, Armidale, N.S.W.

In the usual method there are two tensor equations depending on the two antisymmetrical tensors, each with six independent components, corresponding to the bivectors (**a**). In our notation, the components of these tensors are, respectively,

$$
\begin{pmatrix}
0 & -B_3 & B_2 & iE_1 \\
B_3 & 0 & -B_1 & iE_2 \\
-B_2 & B_1 & 0 & iE_3 \\
-iE_1 & -iE_2 & -iE_3 & 0
\end{pmatrix}
\text{ and }
\begin{pmatrix}
0 & -H_3 & H_2 & iD_1 \\
H_3 & 0 & -H_1 & iD_2 \\
-H_2 & H_1 & 0 & iD_3 \\
-iD_1 & -iD_2 & -iD_3 & 0
\end{pmatrix}.
$$

.... (**b**)

In this form the equations may be readily extended to the continuum of General Relativity. Examples of this tensor method are given in works by Silberstein (1924, pp. 288, 462), McConnell (1936), Costa de Beauregard (1949), and Schouten (1951). The method is based on the investigations of Minkowski (1910) in Special Relativity. He uses a matrix notation, obtaining two equations (loc. cit., p. 38, equations {A} and {B}) in terms of the two matrices corresponding to (**b**). By simply adding Minkowski's two equations we have the equivalent of the single field equation of the present theory (equation (**1**)), for the special case of Cartesian coordinate systems.

Sommerfeld (1948) mentions two bivectors similar to (**a**), but gives the electromagnetic equations in their usual tensor form.

II. MAXWELL'S EQUATIONS AND THE FIELD COMPONENTS

Consider the equation*

$$
\frac{\partial}{\partial x^\nu}(\mathrm{D}^{\mu\nu}) = \mathrm{J}^\mu, \qquad \cdots\cdots\cdots\cdots\cdots\cdots (1)
$$

where J^μ is the current four-vector density

$$
\mathrm{J}^\mu = -\frac{1}{c}(j_1,\, j_2,\, j_3,\, ic\rho), \qquad \cdots\cdots\cdots\cdots (2)
$$

and $\mathrm{D}^{\mu\nu}$ is the field tensor density derived from the field tensor $D^{\mu\nu}$, whose components in the geodesic Cartesian system $(x,\, y,\, z,\, ict)$ are

$$
D^{\mu\nu} =
\begin{pmatrix}
0 & F_3 & -F_2 & N_1 \\
-F_3 & 0 & F_1 & N_2 \\
F_2 & -F_1 & 0 & N_3 \\
-N_1 & -N_2 & -N_3 & 0
\end{pmatrix},
\qquad \cdots\cdots\cdots\cdots (3)
$$

where

$$
\left.
\begin{array}{l}
F_1 = iE_1 - H_1,\ \text{etc.,} \\
N_1 = iD_1 - B_1,\ \text{etc.}
\end{array}
\right\} \qquad \cdots\cdots\cdots\cdots (4)
$$

It will be noticed that (**3**) may be obtained from (**b**) by combining one with the dual of the other. However, (**4**) are not the components of Silberstein's bivectors (**a**) except in a free medium, the case previously considered.

If (**1**) is written out in the geodesic Cartesian system, we have, from the real part,

$$
\frac{\partial \mathbf{D}}{\partial t} + \mathbf{j} = c\ \mathrm{curl}\ \mathbf{H},
$$

$$
\mathrm{div}\ \mathbf{B} = 0,
$$

---

* Roman type shall denote tensor densities, while italics shall refer to tensors.

and from the imaginary part,

$$\frac{\partial \mathbf{B}}{\partial t} = -c \text{ curl } \mathbf{E}$$

$$\text{div } \mathbf{D} = \rho.$$

Equations (1), (2), (3), and (4) therefore define the complete set of Maxwell's equations in macroscopic terms, for a material medium at rest relative to the observer. Instead of two sets of equations depending on two field tensors, each with six terms, we now have one equation depending on one tensor with six distinct components.

If it is assumed that (3) and (4) determine the classical components of the field in *any* system of coordinates, an inconsistent set of equations is obtained for the transformation of the field components. We must therefore postulate some other tensor, having fewer distinct terms, whose transformation is to define the field in all systems of coordinates. This tensor shall be, in the geodesic Cartesian system,

$$E^{\mu\nu} = \left.\begin{array}{cccc} 0 & M_3 & -M_2 & M_1 \\ -M_3 & 0 & M_1 & M_2 \\ M_2 & -M_1 & 0 & M_3 \\ -M_1 & -M_2 & -M_3 & 0 \end{array}\right\}, \quad \dots\dots\dots \quad (5)$$

where

$$M_1 = iE_1 - B_1, \text{ etc.} \quad \dots\dots\dots\dots \quad (6)$$

This is constructed from the first of (b), the tensor which defines the transformation of the field components in the usual relativistic theory of the material medium, and it has the same form as the $F^{\mu\nu}$ which was used in the case of fields in free space (Taylor 1952).

The relation of the field components in one Cartesian system of coordinates (x^μ) to the components in another system (x'^μ) is then given by the transformation equations

$$E'^{\mu\nu} = \frac{\partial x'^\mu}{\partial x^\alpha} \frac{\partial x'^\nu}{\partial x^\beta} E^{\alpha\beta},$$

where, according to our postulate we have

$$E'^{12} = M'_1 = iE'_1 - B'_1, \text{ etc.}$$

as well as (5) and (6). Writing this out, using the coefficients $\partial x'^\mu/\partial x^\alpha$ calculated from the Lorentz transformation, we obtain the required classical equations

$$E'_1 = E_1, \qquad\qquad B'_1 = B_1,$$

$$E'_{2,3} = \frac{E_{2,3} \mp (v/c)B_{3,2}}{\sqrt{1 - v^2/c^2}}, \quad B'_{2,3} = \frac{B_{2,3} \pm (v/c)E_{3,2}}{\sqrt{1 - v^2/c^2}},$$

for the case where the space of (x'^μ) is moving relative to the space of (x^μ) with velocity v in the x^1 direction.

The fact that the components of $E^{\mu\nu}$ are arranged as shown in (5) can be stated in the covariant form

$$E^{\mu\nu} = \tfrac{1}{2}\varepsilon^{\mu\nu\sigma\tau}E_{\sigma\tau}. \quad \dots\dots\dots\dots \quad (7)$$

A tensor which satisfies an equation of this type is described by the term " self-dual " since the operation indicated on the right-hand side (that is, the formation

of the "dual") reproduces the original tensor. Guided by (7) we shall try expressing $E\mu\nu$ in terms of a potential function as follows:

$$E\mu\nu = \left(\sqrt{g}g\mu\sigma g\nu\tau + \tfrac{1}{2}\varepsilon\mu\nu\sigma\tau\right)\left(\frac{\partial\varkappa_\sigma}{\partial x^\tau} - \frac{\partial\varkappa_\tau}{\partial x^\sigma}\right). \quad \dots\dots\dots \text{(8)}$$

This identically satisfies (7). By comparing with the similar treatment for $F\mu\nu$ in the case of empty space, we find that (8) includes the classical equations

$$\mathbf{B} = \text{curl } \mathbf{A},$$
$$\mathbf{E} = -\text{grad } V - \dot{\mathbf{A}}/c,$$

where

$$\varkappa_\mu = (A_1,\ A_2,\ A_3,\ iV).$$

We now show that the six-component system $D\mu\nu$ can be expressed in terms of $E\mu\nu$ and another three-component system, $B\mu\nu$. Take, for the case of the geodesic Cartesian system at rest in the medium,

$$B\mu\nu = \begin{matrix} 0 & 0 & 0 & L_1 \\ 0 & 0 & 0 & L_2 \\ 0 & 0 & 0 & L_3 \\ -L_1 & -L_2 & -L_3 & 0, \end{matrix}$$

where

$$L_1 = N_1 - F_1 = i(D_1 - E_1) - (B_1 - H_1), \text{ etc.}$$

This tensor vanishes in an empty region, and is of an appropriate form for matter at rest, since it attributes a special distinction to components involving a time index. It therefore seems a suitable tensor for describing the departure of the field, in the presence of matter, from the completely isotropic (i.e. invariant in form under coordinate transformations) part represented by $E\mu\nu$. We then have

$$D\mu\nu = E\mu\nu + \mathbf{I}B\mu\nu - \tfrac{1}{2}\varepsilon\mu\nu\sigma\tau\mathbf{R}B\sigma\tau, \quad \dots\dots\dots\dots \text{(9)}$$

where the $\mathbf{I}$, $\mathbf{R}$ operators denote that the imaginary and real parts, respectively, are taken, the factor i being preserved in the imaginary part so that $B\mu\nu = (\mathbf{R}+\mathbf{I})B\mu\nu$. For the Cartesian system,

$$\mathbf{I}B\mu\nu = \begin{matrix} 0 & 0 & 0 & i(D_1-E_1) \\ 0 & 0 & 0 & i(D_2-E_2) \\ 0 & 0 & 0 & i(D_3-E_3) \\ -i(D_1-E_1) & -i(D_2-E_2) & -i(D_3-E_3) & 0 \end{matrix},$$

and

$$-\tfrac{1}{2}\varepsilon\mu\nu\sigma\tau\mathbf{R}B\sigma\tau = \begin{matrix} 0 & (B_3-H_3) & -(B_2-H_2) & 0 \\ -(B_3-H_3) & 0 & (B_1-H_1) & 0 \\ (B_2-H_2) & -(B_1-H_1) & 0 & 0 \\ 0 & 0 & 0 & 0 \end{matrix}.$$

Hence (9) implies

$$D^{12} = iE_3 - B_3 + 0 + (B_3 - H_3) = iE_3 - H_3, \text{ etc.}$$
$$D^{14} = iE_1 - B_1 + i(D_1 - E_1) + 0 = iD_1 - B_1, \text{ etc.}$$

as required.

Using (9) the field equations can be expressed by one equation involving two tensors with complex terms, each having three distinct components.

It will be observed that the space-space components of the tensor $D^{\mu\nu}$, given by equation (3), are the components of the self-dual tensor $F^{\mu\nu}$, which has been used in the treatment of fields in free space. Since $F^{\mu\nu}$ is independent of the properties of the medium, and since it is a characteristic of the self-dual part that it does not indicate motion relative to the medium (being preserved in the same form for transformations of coordinates), it might at first seem more reasonable to choose $F^{\mu\nu}$ instead of $E^{\mu\nu}$ as the self-dual portion of $D^{\mu\nu}$. However, if this were done, results would not agree with those of classical theory except when $\mu=1$. Hence $E^{\mu\nu}$, but not $F^{\mu\nu}$, is an isotropic tensor in a material medium in which $\mu\neq1$.

III. Relations between the Field Components

The next step is to express the classical equations

$$\mathbf{D}=\varepsilon\mathbf{E}, \qquad \mathbf{B}=\mu\mathbf{H}$$

for a homogeneous medium, as a single tensor equation. Since there are six different equations involved here, we shall attempt to solve the problem by assuming a completely self-dual set such as

$$D^{\mu\nu}+\tfrac{1}{2}\varepsilon^{\mu\nu\sigma\tau}D_{\sigma\tau}=a\mathbf{R}E^{\mu\nu}+b\mathbf{I}E^{\mu\nu}, \quad \dots\dots\dots (10)$$

the $E^{\mu\nu}$ being split up into its (self-dual) real and imaginary parts to admit the appearance of two constants. The comparison of (10) with the classical equations provides a, b in terms of ε, μ.

When the indices $\mu,\nu=1,2$ or $3,4$ in a Cartesian system at rest in the medium,

$$F_3+N_3=a\mathbf{R}M_3+b\mathbf{I}M_3.$$

Therefore

$$(iE_3-H_3)+(iD_3-B_3)=-aB_3+biE_3,$$

with similar expressions for the other components. The real and imaginary parts of this will respectively reduce to the third components of the classical equations quoted above, provided

$$\left.\begin{array}{l} a=1+1/\mu=1+\varphi,\ \text{say,} \\ b=1+\varepsilon. \end{array}\right\} \quad \dots\dots\dots\dots\dots (11)$$

From (10) and (11) therefore,

$$D^{\mu\nu}+\tfrac{1}{2}\varepsilon^{\mu\nu\sigma\tau}D_{\sigma\tau}=E^{\mu\nu}+\varphi\mathbf{R}E^{\mu\nu}+\varepsilon\mathbf{I}E^{\mu\nu}. \quad \dots\dots\dots (12)$$

Other forms may be obtained by substituting from (9) in this equation, or they may be derived directly by the method just described. For example, commencing with the completely self-dual equation

$$D^{\mu\nu}+\tfrac{1}{2}\varepsilon^{\mu\nu\sigma\tau}D_{\sigma\tau}=\alpha E^{\mu\nu}+\beta(B^{\mu\nu}+\tfrac{1}{2}\varepsilon^{\mu\nu\sigma\tau}B_{\sigma\tau}),$$

we find

$$\alpha=\frac{2(\varepsilon-\mu)}{\mu\varepsilon-2\mu+1}, \qquad \beta=\frac{\mu\varepsilon-1}{\mu\varepsilon-2\mu+1}.$$

In the case of a crystalline medium we require an equation reducing in the rest system to

$$D_i=\varepsilon_{ij}E_j, \qquad B_i=\mu_{ij}H_j,$$

where $i,j=1,2,3$. The second equation may be solved to give

$$H_i=\varphi_{ij}B_j.$$

There is also the additional requirement that the nine-component systems $\varepsilon_{ij}, \varphi_{ij}$ should be symmetrical.

Consider the equation*

$$D^{\mu\nu}+\tfrac{1}{2}\varepsilon^{\mu\nu\sigma\tau}D_{\sigma\tau}-E^{\mu\nu}=\tfrac{1}{4}(\varphi^{\mu\nu}_{\alpha\beta}\mathbf{R}+\varepsilon^{\mu\nu}_{\alpha\beta}\mathbf{I})E^{\alpha\beta}. \quad \ldots\ldots\ldots (13)$$

Since there are only three independent terms in the tensor which constitutes the whole of the left-hand side, and only three in the $E^{\alpha\beta}$, the $\varphi^{\mu\nu}_{\alpha\beta}$ and $\varepsilon^{\mu\nu}_{\alpha\beta}$ must each have only 3×3 independent components, as required.

Using geodesic coordinates at rest in the medium, (13) gives for $\mu,\nu=1,2$; 2,1 ; 3,4 ; or 4,3

$$D^{12}+D^{34}-E^{12}=\tfrac{1}{4}(\varphi^{12}_{\alpha\beta}\mathbf{R}+\varepsilon^{12}_{\alpha\beta}\mathbf{I})E^{\alpha\beta}.$$

In the sums over α,β there are four terms equal to that in which $\alpha,\beta=1,2$. Similarly for $\alpha,\beta=2,3$ and 3,1. Hence

$$D^{12}+D^{34}-E^{12}=(\varphi^{12}_{12}\mathbf{R}+\varepsilon^{12}_{12}\mathbf{I})E^{12}+(\varphi^{12}_{23}\mathbf{R}+\varepsilon^{12}_{23}\mathbf{I})E^{23}+(\varphi^{12}_{31}\mathbf{R}+\varepsilon^{12}_{31}\mathbf{I})E^{31}.$$

Equating real and imaginary parts, this gives

$$H_3=\varphi^{12}_{12}B_3+\varphi^{12}_{31}B_2+\varphi^{12}_{23}B_1,$$
$$D_3=\varepsilon^{12}_{12}E_3+\varepsilon^{12}_{31}E_2+\varepsilon^{12}_{23}E_1.$$

The formulae for H_2, D_2 and H_1, D_1 are found similarly. The results can be summarized by

$$\varepsilon_{ij}=\begin{matrix} \varepsilon^{23}_{23} & \varepsilon^{23}_{31} & \varepsilon^{23}_{12} \\ \varepsilon^{31}_{23} & \varepsilon^{31}_{31} & \varepsilon^{31}_{12} \\ \varepsilon^{12}_{23} & \varepsilon^{12}_{31} & \varepsilon^{12}_{12} \end{matrix} = \left.\begin{matrix} \varepsilon^{14}_{14} & \varepsilon^{14}_{24} & \varepsilon^{14}_{34} \\ \varepsilon^{24}_{14} & \varepsilon^{24}_{24} & \varepsilon^{24}_{34} \\ \varepsilon^{34}_{14} & \varepsilon^{34}_{24} & \varepsilon^{34}_{34} \end{matrix}\right\}, \text{ etc.,}$$

each independent $\varepsilon^{\mu\nu}_{\alpha\beta}$ corresponding to one of the ε_{ij}. Similarly for the φ_{ij}.

The symmetry of ε_{ij} shows that

$$\varepsilon^{\mu\nu}_{\alpha\beta}=\varepsilon^{\alpha\beta}_{\mu\nu},$$

and similarly

$$\varphi^{\mu\nu}_{\alpha\beta}=\varphi^{\alpha\beta}_{\mu\nu}.$$

These two latter equations are not transformable in general coordinate systems.

In the usual theory, different methods of deriving covariant forms of the classical relations $\mathbf{D}=\varepsilon\mathbf{E}$, $\mathbf{B}=\mu\mathbf{H}$ have been used. Schouten (1951) obtains a single tensor equation with six components, while McConnell (1936) takes, for the more general case of the crystalline medium, two tensor equations.

IV. The Ponderomotive Force

Consider

$$W^\mu=\frac{1}{g}D^{\mu\nu}J_\nu=D^{\mu\nu}J_\nu. \quad \ldots\ldots\ldots\ldots (14)$$

Take the case of geodesic Cartesian coordinates. In such a system a force/power four-vector P^μ would be given by

$$P^\mu=\mathbf{R}(W^1,\ W^2,\ W^3),\ \mathbf{I}(W^4).$$

* The dielectric constants must not be confused with the ε-systems, in which the suffixes will always appear as superscripts.

Now

$$P^1 = \mathbf{R}(D^{12}J^2 + D^{13}J^3 + D^{14}J^4)$$

$$= \mathbf{R}\left\{-\frac{1}{c}(F_3 j_2 - F_2 j_3) - N_1 \mathrm{i}\rho\right\}$$

$$= \frac{1}{c}(H_3 j_2 - H_2 j_3) + D_1\rho.$$

Also

$$P^4 = \mathbf{I}(D^{41}J^1 + D^{42}J^2 + D^{43}J^3)$$

$$= \frac{\mathrm{i}}{c}\mathbf{j} \cdot \mathbf{D}.$$

Therefore

$$P^\mu = \left(\rho\mathbf{D} + \frac{1}{c}\mathbf{j}\times\mathbf{H}, \ \frac{\mathrm{i}}{c}\mathbf{j}\cdot\mathbf{D}\right).$$

Hence (**14**) corresponds to the classical force/power per unit volume.

The conservation of charge is deduced exactly as before.

The law of propagation of electromagnetic waves has lost its fundamental significance—the velocity is no longer the same for all observers—and so the wave equations cannot be derived as neatly as before from the general four-dimensional equations. They can best be obtained by way of the classical equations.

V. The Propagation of Potential

We now attempt the deduction of the wave equation for the potential function from the four-dimensional expressions. Consider the case of a homogeneous isotropic medium at rest with respect to a geodesic system of coordinates. Equation (**12**) becomes

$$D^{\mu\nu} + \tfrac{1}{2}\varepsilon^{\mu\nu\sigma\tau}D\sigma\tau - E^{\mu\nu} = (\varphi\mathbf{R} + \varepsilon\mathbf{I})E^{\mu\nu}. \quad \dots\dots\dots (15)$$

Also, from (**9**),

$$D^{\mu\nu} = E^{\mu\nu} + \mathbf{I}B^{\mu\nu} - \tfrac{1}{2}\varepsilon^{\mu\nu\sigma\tau}\mathbf{R}B\sigma\tau. \quad \dots\dots\dots\dots (16)$$

From the definition of the potential function (**8**) and equation (**7**) which it includes,

$$E^{\mu\nu} = \frac{\partial\varkappa_\mu}{\partial x^\nu} - \frac{\partial\varkappa_\nu}{\partial x^\mu} + \tfrac{1}{2}\varepsilon^{\mu\nu\sigma\tau}\left(\frac{\partial\varkappa_\sigma}{\partial x^\tau} - \frac{\partial\varkappa_\tau}{\partial x^\sigma}\right), \quad \dots\dots\dots\dots (17)$$

$$E^{\mu\nu} = \tfrac{1}{2}\varepsilon^{\mu\nu\sigma\tau}E\sigma\tau. \quad \dots\dots\dots\dots\dots\dots\dots (17')$$

From (**16**) and (**17'**)

$$\tfrac{1}{2}\varepsilon^{\mu\nu\sigma\tau}D\sigma\tau = E^{\mu\nu} + \tfrac{1}{2}\varepsilon^{\mu\nu\sigma\tau}\mathbf{I}B\sigma\tau - \mathbf{R}B^{\mu\nu}. \quad \dots\dots\dots (18)$$

Substituting from (**18**) in (**15**),

$$D^{\mu\nu} + \tfrac{1}{2}\varepsilon^{\mu\nu\sigma\tau}\mathbf{I}B\sigma\tau - \mathbf{R}B^{\mu\nu} - (\varphi\mathbf{R} + \varepsilon\mathbf{I})E^{\mu\nu} = 0. \quad \dots\dots (19)$$

We shall let Greek suffixes range from 1 to 4, and Latin suffixes from 1 to 3. In (**19**) we take $\mu,\nu = 4,i$. Then in the second term σ,τ must be of the type j,k. But $B^{jk} = 0$. Hence

$$D^{4i} - \mathbf{R}B^{4i} - \varphi\mathbf{R}E^{4i} - \varepsilon\mathbf{I}E^{4i} = 0. \quad \dots\dots\dots\dots (20)$$

Assuming that there are no currents or free charges, we have, from (**1**)

$$\partial D^{4i}/\partial x^i = 0,$$

and so, operating on (20) with $\partial/\partial x^i$, and choosing the imaginary part,

$$-\varepsilon\mathbf{l}\frac{\partial E^{4i}}{\partial x^i}=0. \quad \dots\dots\dots\dots\dots\dots \quad (21)$$

But, from (17)

$$E^{4i}=\frac{\partial\varkappa_4}{\partial x^i}-\frac{\partial\varkappa_i}{\partial x^4}+\tfrac{1}{2}\varepsilon^{4ijk}\left(\frac{\partial\varkappa_j}{\partial x^k}-\frac{\partial\varkappa_k}{\partial x^j}\right).$$

Hence

$$\frac{\partial E^{4i}}{\partial x^i}=\frac{\partial}{\partial x^i}\left(\frac{\partial\varkappa_4}{\partial x^i}-\frac{\partial\varkappa_i}{\partial x^4}\right).$$

For a reason which will appear later we cannot here use $\partial\varkappa_\mu/\partial x^\mu=0$, as in the case of empty space. Since $\varkappa_4$ and x^4 are both imaginary, (21) becomes

$$\frac{\partial}{\partial x^i}\left(\frac{\partial\varkappa_4}{\partial x^i}-\frac{\partial\varkappa_i}{\partial x^4}\right)=0.$$

Writing $\varkappa_i=\mathbf{A}$, $\varkappa_4=iV$, this gives

$$\nabla^2 V+\frac{1}{c}\operatorname{div}\dot{\mathbf{A}}=0. \quad \dots\dots\dots\dots\dots\dots \quad (22)$$

Now let us take $\mu,\nu=i,j$ in (15) and (16), noting that $B^{ij}=0$, and also that $\varepsilon^{ij\sigma\tau}$ has the two sets of components $\varepsilon^{ijk4}=-\varepsilon^{ij4k}$. Subtracting the two equations thus obtained in order to eliminate D^{ij},

$$\varepsilon^{ijk4}D^{k4}-(\varphi\mathbf{R}+\varepsilon\mathbf{l})E^{ij}-\varepsilon^{ijk4}\mathbf{R}B^{k4}=0.$$

We next differentiate this with respect to x^4. Now

$$\partial D^{k\nu}/\partial x^\nu=0,$$

and so

$$\partial D^{k4}/\partial x^4=-\partial D^{kn}/\partial x^n.$$

Hence

$$-\varepsilon^{ijk4}\frac{\partial D^{kn}}{\partial x^n}-\frac{\partial}{\partial x^4}(\varphi\mathbf{R}+\varepsilon\mathbf{l})E^{ij}-\varepsilon^{ijk4}\frac{\partial}{\partial x^4}(\mathbf{R}B^{k4})=0.$$

From (19), using $B^{kn}=0$ again,

$$D^{kn}=-\varepsilon^{knp4}\mathbf{l}B^{p4}+(\varphi\mathbf{R}+\varepsilon\mathbf{l})E^{kn}.$$

Hence

$$-\varepsilon^{ijk4}\frac{\partial}{\partial x^n}\{-\varepsilon^{knp4}\mathbf{l}B^{p4}+(\varphi\mathbf{R}+\varepsilon\mathbf{l})E^{kn}\}-\frac{\partial}{\partial x^4}(\varphi\mathbf{R}+\varepsilon\mathbf{l})E^{ij}-\varepsilon^{ijk4}\frac{\partial}{\partial x^4}(\mathbf{R}B^{k4})=0.$$

The real part of this is obviously

$$-\varepsilon^{ijk4}\frac{\partial}{\partial x^n}(\varphi\mathbf{R}E^{kn})-\frac{\partial}{\partial x^4}(\varepsilon\mathbf{l}E^{ij})=0. \quad \dots\dots\dots\dots \quad (23)$$

Now, from (17)

$$E^{ij}=\frac{\partial\varkappa_i}{\partial x^j}-\frac{\partial\varkappa_j}{\partial x^i}+\left(\frac{\partial\varkappa_k}{\partial x^4}-\frac{\partial\varkappa_4}{\partial x^k}\right),$$

where i,j,k is an *even* permutation of 1,2,3. Hence,

$$\mathbf{l}E^{ij}=\left(\frac{\partial\varkappa_k}{\partial x^4}-\frac{\partial\varkappa_4}{\partial x^k}\right),$$

$$\frac{\partial}{\partial x^n}(\mathbf{R}E^{1n})=\frac{\partial}{\partial x^2}\left(\frac{\partial\varkappa_1}{\partial x^2}-\frac{\partial\varkappa_2}{\partial x^1}\right)+\frac{\partial}{\partial x^3}\left(\frac{\partial\varkappa_1}{\partial x^3}-\frac{\partial\varkappa_3}{\partial x^1}\right).$$

Putting $i,j=2,3$ in (23) and substituting from these,

$$-\varphi\left\{\frac{\partial}{\partial x^2}\left(\frac{\partial \varkappa_1}{\partial x^2}-\frac{\partial \varkappa_2}{\partial x^1}\right)+\frac{\partial}{\partial x^3}\left(\frac{\partial \varkappa_1}{\partial x^3}-\frac{\partial \varkappa_3}{\partial x^1}\right)\right\}-\varepsilon\frac{\partial}{\partial x^4}\left(\frac{\partial \varkappa_1}{\partial x^4}-\frac{\partial \varkappa_4}{\partial x^1}\right)=0.$$

Substituting for x^μ and $\varkappa_4$ this is seen to be the x-component of

$$\frac{\varepsilon}{c^2}\ddot{\mathbf{A}}+\frac{\varepsilon}{c}\operatorname{grad}\dot{V}+\varphi\operatorname{grad}\operatorname{div}\mathbf{A}-\varphi\bigtriangledown^2\mathbf{A}=0.$$

Putting $i,j=3,1$ and $1,2$ gives the other components. This can be written

$$\frac{\varepsilon\mu}{c^2}\ddot{\mathbf{A}}-\bigtriangledown^2\mathbf{A}+\operatorname{grad}\left(\operatorname{div}\mathbf{A}+\frac{\varepsilon\mu}{c}\dot{V}\right)=0. \quad \ldots\ldots (24)$$

In order that the propagation of the scalar potential V and the vector potential $\mathbf{A}$ should be consistent, we require

$$\operatorname{div}\mathbf{A}+\frac{\varepsilon\mu}{c}\dot{V}=k, \quad \ldots\ldots\ldots\ldots\ldots\ldots (25)$$

where k is a constant, usually taken to be zero. This causes (24) to be reduced to

$$\bigtriangledown^2\mathbf{A}=\frac{\varepsilon\mu}{c^2}\ddot{\mathbf{A}},$$

and on substituting in (22) gives

$$\bigtriangledown^2 V=\frac{\varepsilon\mu}{c^2}\ddot{V}.$$

The condition (25) shows why it was not possible to assume $\partial \varkappa_\mu/\partial x^\mu=0$, as for empty space. This equation must be replaced by the (covariant) statement that all components of the potential function are propagated with the same velocity.

VI. References

COSTA DE BEAUREGARD, O. (1949).—" La Théorie de la Relativité Restreinte." p. 70. (Masson et Cie : Paris.)

McCONNELL, A. J. (1936).—" Applications of the Absolute Differential Calculus." p. 298. (Blackie & Son Ltd. : London and Glasgow.)

MINKOWSKI, H. (1910).—" Zwei Abhandlungen über die Grundgleichungen der Electrodynamik." p. 38. (B. G. Teubner : Leipzig.)

SCHOUTEN, J. A. (1951).—" Tensor Analysis for Physicists." p. 214. (Oxford Univ. Press.)

SILBERSTEIN, L. (1907).—Nachtrag zur Abhandlung über Electromagnetische Grundgleichungen in bivektorieller Behandlung. *Ann. Phys. Lpz.* **24** : 783. (Note to previous paper.)

SILBERSTEIN, L. (1924).—" The Theory of Relativity." (Macmillan & Co. : London.)

SOMMERFELD, A. (1948).—" Vorlesungen über theoretische Physik." Vol. 3. p. 221. (Dieterich Verlagsbuchhandlung : Wiesbaden.)

TAYLOR, N. W. (1952).—A simplified form of the relativistic electromagnetic equations. *Aust. J. Sci. Res.* A **5** : 423.

THE SURFACE TEMPERATURE OF THE MOON

By J. C. Jaeger*

Summary

Results of numerical calculations are given for (i) the fall in surface temperature during a lunar eclipse, (ii) the variation in surface temperature at the equator during a lunation, and (iii) the variation in microwave temperature at the equator during a lunation. These calculations are made on the assumption that heat is lost from the surface by radiation according to the fourth power law. Two models are considered, firstly, that in which the surface material of the Moon is homogeneous, and secondly, that in which it consists of a thin skin of poor conductor on a better conducting substratum.

The experimental results are discussed in the light of these calculations and it is found that none of the proposed models fits them all adequately, and, though there is a slight preference for the thin skin model over the homogeneous solid, it is not possible to discriminate between the two on the information at present available.

I. Introduction

Observations of the surface temperature of the Moon have been available for many years and attempts have been made to deduce information about the nature of the surface from them. The theoretical background necessary for such attempts is the theory of conduction of heat in the semi-infinite solid with prescribed supply of heat at its surface from the Sun, and with loss of heat from its surface by radiation according to the fourth power law. The accurate fourth power law must be used because of the large range of temperatures involved, and, since this makes the problem non-linear, calculations are difficult and must be carried out numerically. The object of this paper is to give numerical information about the solution of this problem for a fairly wide variety of cases, and also some discussion of the experimental results.

In considering the radiation from the Moon as a whole, each element of its surface may be regarded as a semi-infinite solid with its own thermal properties and angle of incidence of the Sun's rays. To get a strict comparison between theory and experiment it would be necessary to integrate the emission from these over the relevant portion of the disk, but as a first approximation all that can be done is to assume that the restricted portion of the disk observed in the optical experiments will behave as a sort of average semi-infinite solid, and to determine the properties of this material.

With regard to the nature of the surface material, three possibilities are to be recognized. Writing K, ρ, c, and $\varkappa = K/\rho c$ for the thermal conductivity, density, specific heat, and diffusivity of the material, and using c.g.s. units,

calorie, and °K., these are : (i) normal igneous rock for which $(K\rho c)^{-\frac{1}{2}}$ has a value of about 20, (ii) a granular or cellular substance such as pumice or lava gravel for which $(K\rho c)^{-\frac{1}{2}}$ is of the order of 100, or (iii) fine dust for which Wesselink (1948) has shown that, under the conditions of low atmospheric pressure existing on the Moon, $(K\rho c)^{-\frac{1}{2}}$ will be of the order of 1000. It has also been suggested by various authors that the surface layers may not be homogeneous in depth, for example, that a thin layer of dust might lie on the surface of a better conductor such as gravel or solid rock. It will be assumed throughout that the thermal properties of the material are independent of temperature ; in view of our ignorance of the nature of the material it seems hardly worth while introducing the additional complication of variation with temperature : this matter is referred to again in Section II.

Three types of experimental information are available, (i) infra-red observations of the fall in temperature during a lunar eclipse, (ii) infra-red observations of the variations in surface temperature during a lunation, and (iii) observations of the variation in radio-microwave temperature during a lunation. Any proposal about the nature of the surface must be consistent with all three.

II. Eclipse Observations

Pettit and Nicholson (1930) in 1927 and Pettit (1940) in 1939 made observations of the fall in temperature of the surface during a lunar eclipse. Epstein (1929) attempted to calculate the thermal properties of the surface from the results for the 1927 eclipse and obtained a value of 120 for $(K\rho c)^{-\frac{1}{2}}$ from which he concluded that the lunar surface is covered with some substance such as pumice. He used a linear theory (in effect assuming that the surface loses heat at a rate proportional to the fourth power of its initial temperature instead of to the fourth power of its actual temperature), and it was pointed out by both Wesselink (1948) and Jaeger and Harper (1950) that, when the accurate fourth power law of radiation is used, a much higher value is obtained : both these authors give single curves calculated for values of $(K\rho c)^{-\frac{1}{2}}$ of the order of 1000 and compare them with the experimental results. A family of such curves for various values of $(K\rho c)^{-\frac{1}{2}}$ is given in Figure 1 which makes the comparison with experiment clearer.

The problem in conduction of heat proposed by the eclipse experiments may be stated as follows. The homogeneous semi-infinite solid $x > 0$ is initially at constant* temperature v_0 (°K.) and in equilibrium with the solar radiation $E\sigma v_0^4$ absorbed by it, where E is its emissivity and σ is the Stefan-Boltzmann constant. During the penumbral stage of the eclipse, the radiation absorbed from the Sun may be written $E\sigma v_0^4 f(t)$, where $f(t)$ can be calculated from the circumstances of the eclipse and vanishes at the beginning of the umbral phase. During both umbra and penumbra the surface loses heat at the rate $E\sigma v_s^4$,

---

* The initial temperature is in fact that due to the monthly periodic variation of surface temperature ; if this is used instead of the constant value the change is very small.

where v_s is its temperature.   Then if t_0 is the duration of penumbra, and v is the temperature in the solid, the equations to be solved are :

$$\frac{\partial^2 v}{\partial x^2} - \frac{1}{\varkappa}\frac{\partial v}{\partial t} = 0, \qquad x>0,\ t>0, \quad \dotsb \quad (1)$$

$$v = v_0, \qquad x>0,\ t=0, \quad \dotsb \quad (2)$$

with the boundary condition at $x=0$

$$K\frac{\partial v}{\partial x} = E\sigma v^4 - E\sigma v_0{}^4 f(t), \qquad 0<t<t_0, \quad \dotsb \quad (3)$$

$$= E\sigma v^4, \qquad\qquad t>t_0. \quad \dotsb \quad (4)$$

Because of the occurrence of the fourth power in (3) and (4) the problem is non-linear and a simple explicit solution for v cannot be obtained.   Wesselink (1948) and Jaeger and Harper (1950) have made numerical calculations using Schmidt's method which is very well adapted to problems of this type (cf. Jaeger 1950).

In Figure 1 a family of curves showing the ratio of the surface temperature v_s to the initial temperature v_0 as a function of t/t_0 is given : the curves are specified by the single parameter

$$P = \frac{\sigma E v_0{}^3 t_0{}^{\frac{1}{2}}}{(10K\rho c)^{\frac{1}{2}}}, \quad \dotsb \quad (5)$$

and the chosen values of P are 2, 1·5, 1, 0·2, and 0·03, corresponding, for $t_0=74$ min., $v_0=370$ °K., and $E=1$, to values 1370, 1026, 685, 137, and 20 of $(K\rho c)^{-\frac{1}{2}}$.   The time interval used in the Schmidt process was $t_0/20$ or 3·7 min. For $f(t)$ a smooth curve drawn through some figures given by Pettit and Nicholson (1930) has been used ; this is the curve marked " Insolation " in Figure 1 ; the temperature is not very sensitive to the form of $f(t)$ and in fact a linear fall $f(t)=(t_0-t)/t_0$ gives curves very little different from those of Figure 1.

The results of Pettit's 1939 eclipse observations are shown by the crosses in Figure 1 and reasonable agreement is obtained if P lies between 1·5 and 2, that is $(K\rho c)^{-\frac{1}{2}}$ between 1370 and 1030.   The values $P=0·2$ and 0·03 correspond to pumice and bare rock respectively, and it appears that the possibility of any large proportion of the surface being covered with these is quite ruled out.   As remarked above, Wesselink (1948) has shown that values of $(K\rho c)^{-\frac{1}{2}}$ of the order of 1000 are likely for dust under lunar conditions, so it seems probable that the surface, or a very considerable fraction of it, is composed of dust.

The most notable discrepancy between the calculated curves of Figure 1 and the experimental values is that the latter fall more slowly in the umbral phase.   Both Jaeger and Harper (1950) and Lettau (1951) remark that this may be due to an increase of thermal conductivity, either with depth or with temperature.

Variation of conductivity with temperature is an attractive explanation but a difficult one to discuss.   Some calculations have been made with conductivity varying as the cube of the absolute temperature but these yield curves which still fall too rapidly in umbra.

The other explanation, an increase in conductivity with depth, which could easily be provided by a layer of dust on a substratum of rock or pumice, can easily be made to give an adequately slow fall of temperature during umbra.

Further, Piddington and Minnett (1949), in interpreting their microwave results, suggested a model consisting of a thin layer of dust on a better-conducting substratum. Using a figure given by them, Jaeger and Harper (1950) give a family of curves for various possible surface layers and get a best fit with 2 mm. of dust on a substratum with $(K\rho c)^{-\frac{1}{2}}$ of the order of 100.
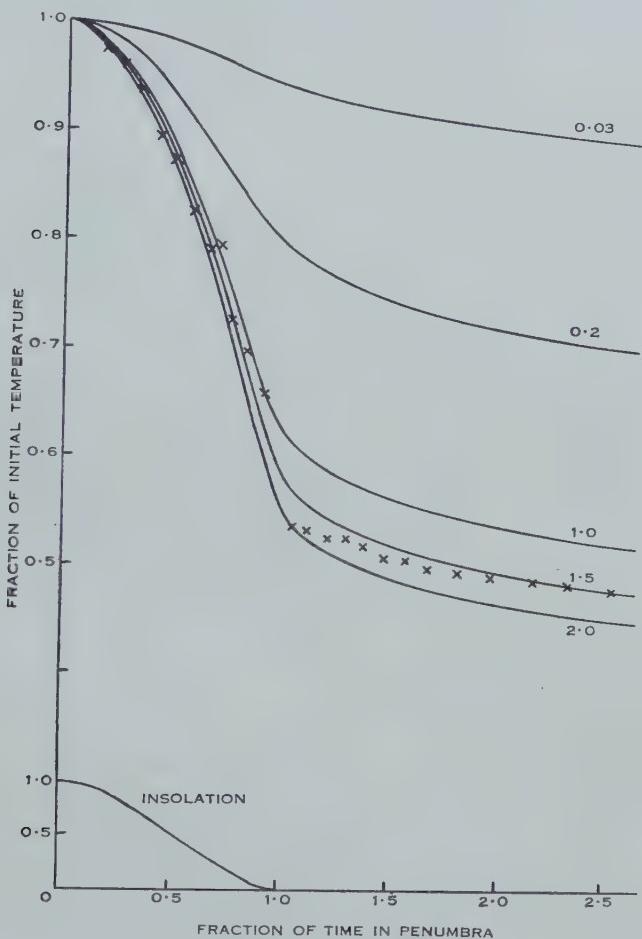
Fig. 1.—Fall in temperature during a lunar eclipse. The numbers on the curves are values of the parameter P. The crosses are the experimental values of Pettit.

To sum up, the eclipse observations at present point definitely to a surface with $(K\rho c)^{-\frac{1}{2}}$ of the order of 1000, which may be interpreted as dust, and suggest that there may only be a few millimetres of this overlying a better conductor. Eclipse observations are much the most satisfactory for a comparison of theory and experiment and it is to be hoped that more will be made so that it may be possible to settle some of the questions raised above.

III. The Variation of the Surface Temperature at the Equator during a Lunation

This has been observed by Pettit and Nicholson (1930) with results which are rather disappointing from the point of view of theoretical interpretation. During the lunar day the temperature reaches a maximum of 374 °K. and is very nearly in equilibrium with the insolation. During the lunar night observations are very difficult and Pettit and Nicholson give the single figure of 120 ± 15 °K. at lunar midnight.
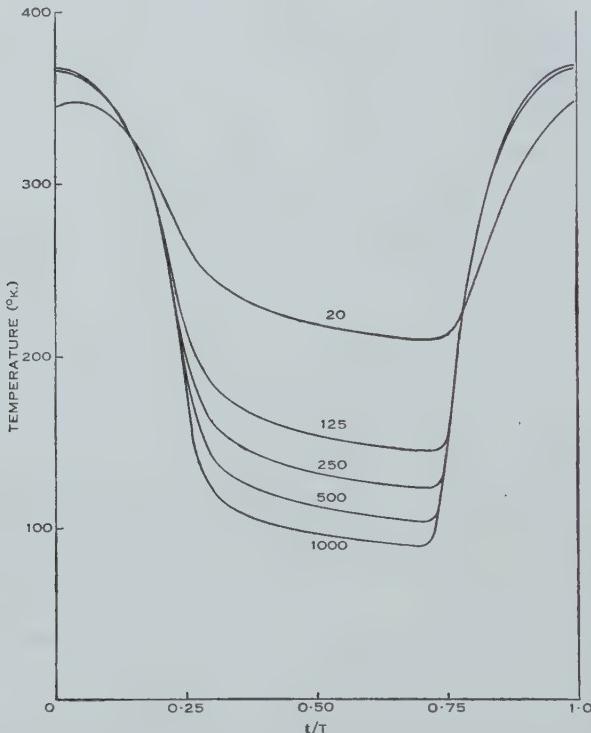
Fig. 2.—Surface temperature of the homogeneous solid. The numbers on the curves are the values of $(K\rho c)^{-\frac{1}{2}}$.

Wesselink (1948) has given a single curve, calculated for $(K\rho c)^{-\frac{1}{2}}=920$ by the Schmidt method, which is in reasonable agreement with Pettit and Nicholson's results. This method is not very suitable for this problem, and Jaeger (1953) has developed a method which permits fairly rapid calculations for periodic problems of this type ; this method, with 20 intervals, has been used for all the calculations described subsequently.

The problem now is to find the steady periodic solution of period T of the equation of conduction of heat (1) for the semi-infinite solid $x>0$ with boundary conditions at $x=0$

$$K\frac{\partial v}{\partial x}=E\sigma v^4-A\ \cos\ \frac{2\pi t}{T}, \quad |\,t\,|<\frac{1}{4}T, \quad \ldots\ldots\ldots\ldots (6)$$

$$=E\sigma v^4, \qquad\qquad \frac{1}{4}T <t<\frac{3}{4}T, \quad \ldots\ldots\ldots (7)$$

where A is the amplitude of the insolation. In all the calculations below we shall take $E=1$; this is justifiable since the emissivity of the lunar surface is known to be high, and reducing the value of E to $0 \cdot 9$ is found to make a change of at most one or two degrees in temperature. The value $0 \cdot 0258$ cal. cm.$^{-2}$ sec.$^{-1}$ will be used for A ; this quantity again is not well known and varies with position in the orbit, but a 10 per cent. change in A gives a change of only one or two degrees in temperature at lunar midnight.
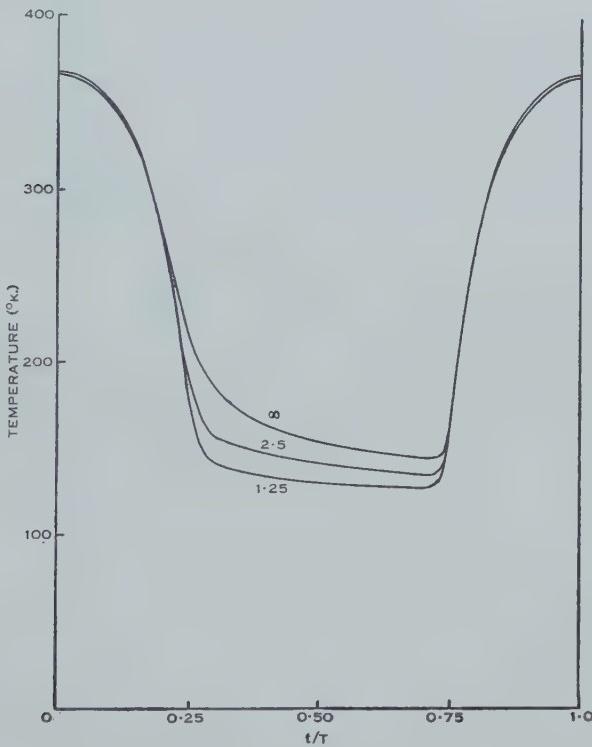
Fig. 3.—Surface temperature of a solid for which $(K\rho c)^{-\frac{1}{2}}=125$, covered by a skin of poor conductor. The numbers on the curves are the values of D specifying the skin.

Figure 2 shows the surface temperature of homogeneous semi-infinite solids with the boundary conditions (6) and (7) and with the values 1000, 500, 250, 125, and 20 of $(K\rho c)^{-\frac{1}{2}}$. Taking 120 °K. as the temperature at lunar midnight, and having regard to the large uncertainty in this value, it appears that values of $(K\rho c)^{-\frac{1}{2}}$ between 200 and 1000 are possible and in particular, as remarked by Wesselink, that dust with $(K\rho c)^{-\frac{1}{2}}$ of the order of 1000 will satisfy both the eclipse observations and the present ones reasonably well.

Next, in view of the suggestion from the eclipse and microwave observations that a two layer model consisting of a thin skin of poor conductor on a better-conducting substratum may fit the results better, it is desirable to make calculations of the surface temperature with this model. Since the skin is thin its

B

thermal capacity may be neglected* for these long period processes, and the model becomes that of a skin of thermal resistance R on a substratum with thermal properties K, ρ, c. The situation may be described in terms of two parameters, $(K\rho c)^{-\frac{1}{2}}$, and

$$D = T^{\frac{1}{2}}(K\rho c)^{-\frac{1}{2}}/R. \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots \quad (8)$$

Since the results are most interesting if the substratum is a much better conductor than the skin (which we shall assume in the calculations to be dust), only the values 20 and 125 for $(K\rho c)^{-\frac{1}{2}}$ will be considered. Also, since Piddington and Minnett (1949) have deduced the value $D=\sqrt{2\pi}$ from their microwave observations, the results have been calculated for this value and simple fractions of it. The results are shown in Figures 3 and 4 for $(K\rho c)^{-\frac{1}{2}}=125$ and 20, respectively. The curves for $D=\infty$ are those of Figure 2 for the homogeneous solid with no surface skin. It appears that the effect of decreasing D is to lower the night-time curves and to flatten them, an effect similar to that remarked in the eclipse calculation. It would be most interesting to have complete experimental curves during the lunar night for comparison.

It appears that "midnight" temperature as low as 120 °K. can be attained with suitable values of D. For solids with $(K\rho c)^{-\frac{1}{2}}$ of 20, 125, and 250 the necessary values of D would be of the orders of $0\cdot1$, $0\cdot6$, and 2 respectively. To see the orders of magnitude involved, suppose the skin has conductivity K' and thickness d, so that $R=d/K'$. If we assume the skin to be dust with the reasonable values $K'=2\cdot8\times10^{-6}$, $\rho'=1\cdot8$, $c'=0\cdot2$, $(K'\rho'c')^{-\frac{1}{2}}\doteqdot1000$, the values of its thickness d for various values of R and $(K\rho c)^{-\frac{1}{2}}$ are shown in Table 1.

TABLE 1

THICKNESS OF DUST CORRESPONDING TO VARIOUS VALUES OF D AND $(K\rho c)^{-\frac{1}{2}}$

D \ $(K\rho c)^{-\frac{1}{2}}$	250	125	20
$2\cdot5$	$0\cdot45$	$0\cdot22$	$0\cdot04$
$1\cdot25$	$0\cdot89$	$0\cdot45$	$0\cdot07$
$0\cdot625$		$0\cdot89$	$0\cdot14$
$0\cdot25$			$0\cdot36$
$0\cdot1$			$0\cdot89$

It appears that in all cases skins of thickness less than 1 cm. are involved.

It may be concluded that, as for the eclipse observations, the results are reasonably well fitted either by a homogeneous solid for which $(K\rho c)^{-\frac{1}{2}}=1000$ or by a layer of such a solid some millimetres thick on a better-conducting substratum.

* It should perhaps be remarked that the thermal capacity of the skin is not neglected in the eclipse calculations ; in fact much of the heat extracted in the penumbral phase comes from the skin.

IV. Radio-Microwave Observations

Rock material is partially transparent to radio microwaves whose wavelengths are of the order of 1 cm. Thus the radiation observed from a solid on these wavelengths is not determined merely by the surface temperature but is the total effect of emission from a region near the surface. If v is the temperature at depth x below the surface of the semi-infinite solid, we shall call

$$v_m = \frac{1}{\alpha} \int_0^\infty e^{-\alpha x} v \, dx \quad \dots\dots\dots\dots\dots\dots (9)$$

the microwave temperature at normal incidence. Here α is an attenuation coefficient characteristic of the material and depending on its electrical con-
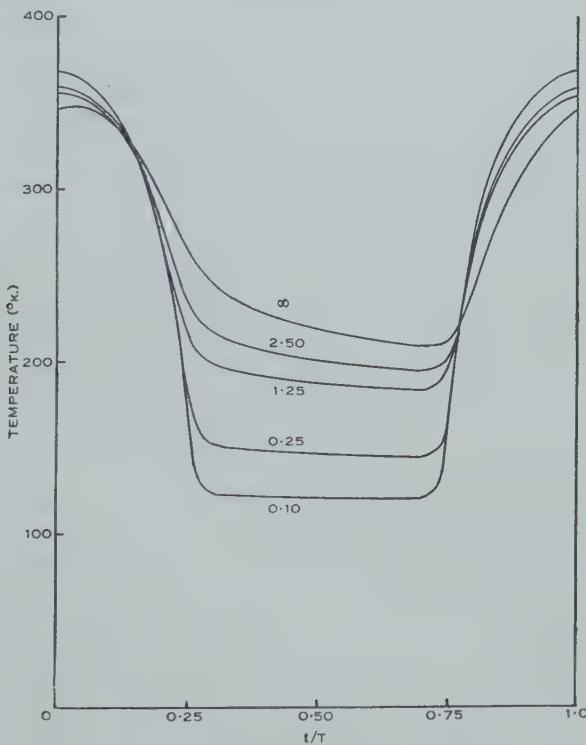
Fig. 4.—Surface temperature of a solid for which $(K\rho c)^{-\frac{1}{2}} = 20$, covered by a skin of poor conductor. The numbers on the curves are the values of D specifying the skin.

ductivity. If the surface is seen from a direction included at θ to the normal, α is to be replaced by $\alpha \sec \theta$. The whole question is discussed in detail by Piddington and Minnett (1949).

Formulae for the calculation of microwave temperatures are given by Jaeger (1953). They involve two parameters, $(K\rho c)^{-\frac{1}{2}}$ and

$$C = \alpha(\varkappa T)^{\frac{1}{2}}. \quad \dots\dots\dots\dots\dots\dots (10)$$

Microwave temperatures at normal incidence, for the homogeneous semi-infinite solids for which $(K\rho c)^{-\frac{1}{2}}$ has the values 1000 and 125, and with the

boundary conditions (**6**) and (**7**), that is, for the Moon at its equator, are shown in Figures 5 and 6 respectively. It appears that decreasing C reduces the amplitude of the oscillation and increases the phase lag of the maximum. $C = \infty$ gives the surface temperature as in Figure 2.

Microwave temperatures may also be calculated for the case of a homogeneous solid covered with a thin skin of poor conductor of negligible heat capacity and from which the microwave emission may be neglected. In this case three parameters are involved, viz. $(K\rho c)^{-\frac{1}{2}}$, D, and C, so that it is
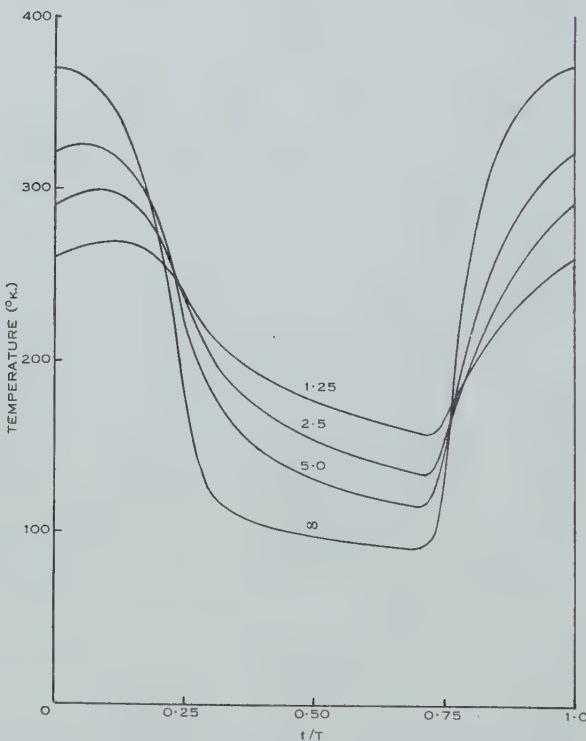
Fig. 5.—Microwave temperatures for a homogeneous solid
for which $(K\rho c)^{-\frac{1}{2}} = 1000$. The numbers on the curves are
the values of $C = \alpha(\varkappa T)^{\frac{1}{2}}$.

impossible to give complete results, but the values of the maximum, mean, and minimum temperatures for a number of cases are shown in Table 2, while curves for the case

$$D = C = \sqrt{2\pi}, \quad \dots \dots \dots \dots \dots \quad (11)$$

which is of special interest in connection with the work of Piddington and Minnett, are given in Figure 7.

The comparison of the experimental microwave results with theory is extremely difficult. It is possible only to observe the whole disk, and not a relatively small area of it as in the optical case : further, the sensitivity of the

equipment varies with position in the aerial beam. Thus the observed result is an average over all latitudes and longitudes, allowing for the angle of emergence of the radiation and position in the aerial beam. Piddington and Minnett

TABLE 2

MAXIMUM, MEAN, AND MINIMUM MICROWAVE TEMPERATURES FOR VARIOUS VALUES OF THE PARAMETERS C AND D

D	C	$(K\rho c)^{-\frac{1}{2}}=1000$			$(K\rho c)^{-\frac{1}{2}}=125$			$(K\rho c)^{-\frac{1}{2}}=20$		
		Max.	Mean	Min.	Max.	Mean	Min.	Max.	Mean.	Min.
∞	∞	370	211	89	366	237	144	347	264	207
	5·0	325	211	114	329	237	164	322	264	221
	2·5	298	211	134	306	237	179	308	264	229
	1·25	269	211	157	283	237	197	292	264	240
2·5	∞				304	231	170	308	257	218
	5·0				284	231	185	294	257	227
	2·5				271	231	195	284	257	233
	1·25				257	231	206	275	257	241
1·25	∞				278	226	181	289	251	221
	5·0				263	226	193	278	251	229
	2·5				254	226	200	272	251	233
	1·25				244	226	208	265	251	239
0·25	∞							249	232	215
	5·0							244	232	220
	2·5							241	232	223
	1·25							238	232	226
0·1	∞							227	219	210
	5·0							225	219	213
	2·5							223	219	214
	1·25							222	219	216

represent their results for the average temperature over the disk by the sinusoid

$$239 + 40\cdot3 \cos\left(\frac{2\pi t}{T} - \frac{\pi}{4}\right) \quad \dots\dots\dots\dots\dots \text{(12)}$$

From this they deduce

$$249 + 52\cdot0 \cos\left(\frac{2\pi t}{T} - \frac{\pi}{4}\right) \quad \dots\dots\dots\dots\dots \text{(13)}$$

for the microwave temperature at normal incidence at the equator (that is, the quantity calculated above). Clearly, the derivation of (13) from (12) involves a great deal of approximation. Finally, they conclude from the phase lag of 45° in (13) that the surface layer of the Moon cannot be regarded as homogeneous but can be represented fairly well by the thin skin model mentioned above and subject to the conditions (11).
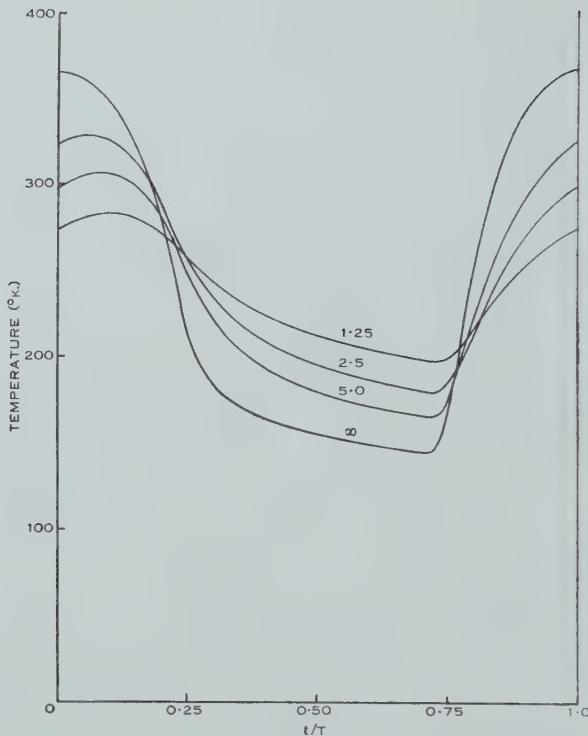
Fig. 6.—Microwave temperatures for a homogeneous solid
for which $(K\rho c)^{-\frac{1}{2}}=125$.   The numbers on the curves are
the values of $C=\alpha(\kappa T)^{\frac{1}{2}}$.
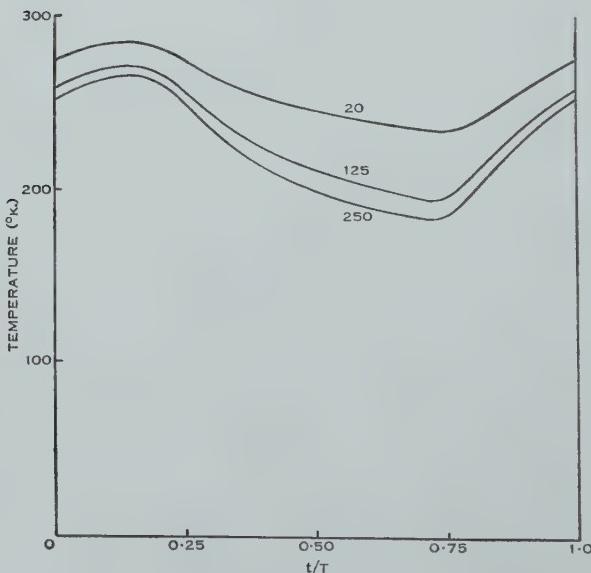
Fig. 7.—Microwave temperatures for a solid with a surface
skin for which $D=C=\sqrt{2\pi}$.   The numbers on the curves
are the values of $(K\rho c)^{-\frac{1}{2}}$.

Their discussion has been criticized by Bracewell (personal communication) who has shown by a more refined analysis of the experimental results (involving both phase and amplitude) that, while a best fit is obtained with values near to (11), a wide range of variation in these parameters is consistent with the experimental results and in particular that the homogeneous solid is not altogether ruled out.

Comparing (13) with the results of Table 2 and Figure 7, it appears that it is difficult to fit both the optical and microwave results with any model. The discrepancy is essentially between the rather high mean temperature demanded by the microwave results and the rather low temperature during the lunar night which the models of Section III were designed to fit. If either of these is relaxed somewhat, reasonable agreement can be obtained with either a homogeneous solid with $(K\rho c)^{-\frac{1}{2}}$ in the range 500 to 1000 or with a thin layer of such a solid on a substratum with $(K\rho c)^{-\frac{1}{2}}$ of the order of 100 ; the third possibility, that of a thin layer on a substratum of rock with $(K\rho c)^{-\frac{1}{2}}$ of the order of 20, seems less likely because it leads to low values of the amplitude of the oscillation of the microwave temperature.

V. ACKNOWLEDGMENTS

VI. REFERENCES

EPSTEIN, P. (1929).—*Phys. Rev.* **33** : 269.
JAEGER, J. C. (1950).—*Proc. Camb. Phil. Soc.* **46** : 634.
JAEGER, J. C. (1953).—*Proc. Camb. Phil. Soc.* (in press).
JAEGER, J. C., and HARPER, A. F. A. (1950).—*Nature,* **166** : 1026.
LETTAU, H. (1951).—*Geofis. Pur. Appl.* **19** : 1.
PETTIT, E. (1940).—*Astrophys. J.* **91** : 408.
PETTIT, E., and NICHOLSON, S. B. (1930).—*Astrophys. J.* **71** : 102.
PIDDINGTON, J. H., and MINNETT, H. C. (1949).—*Aust. J. Sci. Res.* A **2** : 63.
WESSELINK, A. J. (1948).—*Bull. Astr. Insts. Netherlds.* **10** : 351.

EMISSION OF RADIATION FROM MODEL HYDROGEN CHROMOSPHERES

By J. T. Jefferies*

Summary

On the basis of earlier work of Giovanelli (1949), the characteristics of the radiation fields of Hα, Lα, Lβ, and the Lyman continuum have been calculated for model solar atmospheres consisting of a photosphere and an overlying chromosphere which scatters coherently, and which is isothermal at one of a number of kinetic temperatures in the range 1×10^4 to $2 \cdot 5 \times 10^5$ °K.

The computed central intensities (of Hα) suggest that the kinetic temperature of those regions responsible for the observed solar Hα radiation lies somewhere in the range up to about $3 \cdot 5 \times 10^4$ °K. The best agreement with observation, for those temperatures considered, is obtained for $2 \cdot 5 \times 10^4$ °K., with an electron concentration 2×10^{11} per cm.³ at the base of the chromosphere. Too much reliance should not be placed on this result, for the obvious non-uniformity of the solar chromosphere and its effect on the hydrogen lines has not been taken into account.

The effective black-body temperature for the centres of the Lα and Lβ lines and for the Lyman series limit have been calculated ; the results indicate that the intensities in each case pass through a maximum at a kinetic temperature about 5×10^4 °K. The half-widths of Lα and Lβ show a steady decrease with increasing temperature up to 5×10^4 °K. ; for higher temperatures the half-width increases.

The contours of Lα and Hα have also been computed on the basis of a simple model of non-coherent scattering in which the absorbed radiation is redistributed over a Doppler profile ; the results indicate that the type of scattering is unimportant for Hα, but may considerably modify the profiles of the Lyman lines.

I. INTRODUCTION

The discovery that the temperature in the Sun's atmosphere increases outwards has stimulated studies of the emission of radiation from hot atmospheres, such, for example, as those of Thomas (1948, 1949), Giovanelli (1949), and Miyamoto (1951a, 1951b). There is considerable uncertainty, however, as to the temperature distribution in the chromosphere, and to assist studies on this subject, and on the temperatures of disturbed solar regions such as flares and prominences, computations are given here of the radiation emitted by model hydrogen chromospheres over a temperature range $2 \cdot 5 \times 10^5$ to 10^4 °K. Below 10^4 °K. collisions between neutral atoms become relatively important, but as there are no data for the excitation cross sections for these collisions, it has not been possible to extend the calculations to lower temperatures.

Investigations by Thomas (1948, 1949) and Giovanelli (1949) on the intensity of Lyman and Balmer radiation in the solar chromosphere, the former assuming a temperature of 35,000 °K., the latter 25,000 °K., have yielded results in

* Division of Physics, C.S.I.R.O., University Grounds, Sydney.

substantial agreement. The methods adopted by the two authors were rather similar, and involved calculating the equilibrium populations established in a hydrogen atmosphere when transitions between atomic states take place as a result of electron collision or the emission or absorption of radiation. Using an equation of radiative transfer, atomic populations and radiation intensities were obtained.

More recently Miyamoto (1951a, 1951b) has dealt with the excitation of hydrogen and helium in a 30,000 °K. chromosphere. The results obtained for hydrogen are generally in quantitative agreement with those obtained here.

The present method follows closely that of Giovanelli, although improved collision cross sections have been used. The distribution of intensity in the Lyman continuous spectrum and the profiles of Lα, Lβ, and Hα are obtained for chromospheres having a scale height either defined by the observed chromospheric electron gradient, or appropriate to hydrostatic equilibrium. Electron concentrations assumed for the base of the chromosphere range in general from 10^{11} to 10^{12} cm.$^{-3}$. Results are obtained first for an atmosphere which scatters coherently ; non-coherent scattering is considered in Section IX.

II. List of Principal Symbols

The more frequently occurring symbols are as follows :

T, kinetic temperature,

z, vertical height above the base of the atmosphere,

ε_ν, α_ν, " true " emission and absorption coefficients for radiation of frequency ν,

B_ν, defined as E_ν/α_ν, E_ν being the total emission coefficient,

λ_ν, λ'_ν, scattering parameters in the chromosphere and photosphere respectively,

J_ν, total monochromatic intensity,

N_l, concentration of hydrogen atoms in the lth substate,

N_e, N_+, electron and proton concentrations,

N_0, N_e at the base,

β, electron density gradient in $N_e = N_0 \exp(-\beta z)$,

$\Lambda_\alpha N_1$, rate of absorption of Lα radiation per unit volume—similar symbols are used for Lβ and Lyman continuum absorption,

τ_ν, optical depth for the appropriate radiation,

τ_1, τ_1°, values of τ at the base of the atmosphere ; at an arbitrary frequency and at the centre of the line (or series limit) respectively.

III. The Equilibrium Equations

The outward diffusion of radiation from an atmosphere which scatters coherently is given by the approximate relation (see Giovanelli 1949)

$$J_\nu = 4\pi\varepsilon_\nu/\alpha_\nu\lambda_\nu + a_\nu \exp(\sqrt{3\lambda_\nu}\tau_\nu) + b_\nu \exp(-\sqrt{3\lambda_\nu}\tau_\nu), \quad .. \ (1)$$

where J_ν is the total intensity of the radiation ; ε_ν and α_ν are respectively the emission and absorption coefficients ; λ_ν is a scattering parameter, and is the probability that an atom which undergoes an upward transition under the

influence of radiation will be ionized, or return to the ground state by some process not involving the emission of radiation of frequency ν ; and τ_ν is the optical depth defined by $\tau_\nu = \int_\infty^z \alpha_\nu dz$. This solution of the equation of radiative transfer assumes that λ is constant and ε/α is either constant or varies linearly with optical depth. To satisfy these restrictions and simplify the calculations we consider the radiation emitted by an atmosphere of uniform electron concentration and thickness equal to the scale height.

The values of ε, α, and λ depend on the populations of the various atomic states and on the kinetic temperature of the atoms, which we assume equal to the electron temperature.

The populations may be obtained from the condition that, at equilibrium, the rate at which atoms arrive in any quantum state is equal to the rate at which they leave. This may be expressed by the following general equation for the jth substate :

$$\sum_l [A_{l,j} + \beta_{l,j} N_e + I(\nu_{l,j})] N_l + \gamma_j N_e^2 + \delta_j N_e^3$$

$$= \sum_k [A'_{j,k} + \beta'_{j,k} N_e + I'(\nu_{l,k}) + I'(\nu_{j,i}) + \varepsilon_{j,i} N_e] N_j, \quad \ldots\ldots \; (2)$$

where N_l and N_j are respectively the concentrations of atoms in the l and j substates ; N_e is the electron concentration, assumed equal to the proton concentration ; $A_{l,j} N_l$ is the rate of the spontaneous atomic transition $l \to j$; $\beta_{l,j} N_e N_l$ is the rate of the collision excitation or de-excitation for the transition $l \to j$, and $I(\nu_{j,l}) N_l$ is the rate of the same transition under the influence of radiation. The terms on the right of equation (2) are the corresponding transition rates for processes which remove atoms from the j substate. The last two terms on the left of equation (2) represent the rates of spontaneous and three-body recombination ; corresponding terms on the right represent the rates of ionization from the j substate due to absorption of radiation and to electron collision respectively. The rate of the transition $l \to j$ induced by radiation, denoted by $I(\nu_{l,j}) N_l$, is given by

$$I(\nu_{l,j}) N_l = \int_0^\infty \frac{J_\nu \alpha_\nu d\nu}{h\nu}.$$

The set of simultaneous equations represented by (1) and (2) cannot be solved exactly. Following Giovanelli (1949) we shall restrict consideration to the $1S$, $2S$, $2P$, $3S$, $3P$, $3D$, and ionized states, and in equations (2) put the intensities of Hα and of the Balmer and Paschen continua in the chromosphere equal to those in a black body at 5000 °K., assumptions which do not greatly influence the results obtained.

IV. THE TRANSITION RATES

The rates of radiative transition between two substates of the hydrogen atom, i.e. the $A_{l,j}$'s in equation (2), are well known (see, e.g. Unsöld 1938).

The terms $\beta_{l,j}$ and $\beta'_{j,k}$, which represent collision excitations, may be calculated from the appropriate excitation cross sections as functions of the energy of the exciting electrons. Values of the cross sections for excitation from

the $1S$ and $2P$ states for a number of energies, calculated on the assumption of Born's approximation, were kindly made available by Dr. D. R. Bates, of University College, London.   Some of these have subsequently been published (Bates *et al.* 1950).

TABLE 1

RATES OF COLLISION EXCITATION

$(10^{-9} N_n N_e \ \text{cm.}^{-3} \ \text{sec.}^{-1})$

T (°K.)	$1S \to 2S$	$1S \to 2P$	$1S \to 3S$	$1S \to 3P$	$1S \to 3D$	$2P \to 3S$	$2P \to 3P$	$2P \to 3D$
$1\cdot0\times10^4$	$1\cdot08\times10^{-4}$	$3\cdot82\times10^{-4}$	$2\cdot51\times10^{-6}$	$9\cdot14\times10^{-6}$	$9\cdot90\times10^{-7}$	$18\cdot3$	$27\cdot4$	191
$1\cdot5\times10^4$	$6\cdot01\times10^{-3}$	$2\cdot21\times10^{-2}$	$2\cdot86\times10^{-4}$	$1\cdot08\times10^{-3}$	$1\cdot16\times10^{-4}$	$35\cdot3$	$53\cdot7$	417
$2\cdot5\times10^4$	$1\cdot46\times10^{-1}$	$5\cdot83\times10^{-1}$	$1\cdot28\times10^{-2}$	$5\cdot10\times10^{-2}$	$5\cdot40\times10^{-3}$	$56\cdot9$	$85\cdot6$	773
$5\cdot0\times10^4$	$1\cdot50$	$7\cdot06$	$2\cdot01\times10^{-1}$	$8\cdot90\times10^{-1}$	$9\cdot14\times10^{-2}$	$77\cdot6$	107	1210
$1\cdot0\times10^5$	$4\cdot40$	$25\cdot4$	$7\cdot26\times10^{-1}$	$3\cdot83$	$3\cdot69\times10^{-1}$	$86\cdot0$	101	1450
$2\cdot5\times10^5$	$6\cdot42$	$46\cdot8$	$1\cdot20$	$7\cdot64$	$6\cdot76\times10^{-1}$	$80\cdot0$	$74\cdot4$	1385

The rates of electron collision excitation have been found from these values by numerical integration assuming a Maxwellian energy distribution for the electrons at an electron temperature T, results being given in Table 1.   Rates of collision ionization and spontaneous recombination are given in Tables 2 and 3.

TABLE 2

RATES OF COLLISION IONIZATION

$(10^{-7} N_n N_e \ \text{cm.}^{-3} \ \text{sec.}^{-1})$

T (°K.)	$1S \to i$	$2S, 2P \to i$	$3S, 3P, 3D \to i$
$1\cdot0\times10^4$	$6\cdot41\times10^{-8}$	$8\cdot70\times10^{-2}$	$2\cdot86$
$1\cdot5\times10^4$	$1\cdot38\times10^{-5}$	$3\cdot83\times10^{-1}$	$5\cdot40$
$2\cdot5\times10^4$	$1\cdot09\times10^{-3}$	$1\cdot15$	$8\cdot49$
$5\cdot0\times10^4$	$2\cdot86\times10^{-2}$	$2\cdot50$	$1\cdot11\times10$
$1\cdot0\times10^5$	$1\cdot46\times10^{-1}$	$3\cdot17$	$1\cdot13\times10$
$2\cdot5\times10^5$	$3\cdot68\times10^{-1}$	$3\cdot48$	$9\cdot27$

It is worth noting that because of corresponding changes in the rates of superelastic collisions, uncertainties in cross-section data have in many cases little effect on the computed populations and, therefore, on the emergent radiation.

Inelastic collisions resulting in transitions between states of very small energy separation as, for example, $2S \to 2P$ are of some interest.   Purcell (1952) has recently calculated the cross section of this transition using an impact parameter method finding that proton collisions are considerably more important than electron collisions, the effective cross section being of the order of $10^7 \pi a_0^2$. Such a high cross section would greatly simplify our calculations since the $2S$ and $2P$ states would be populated in the ratio of their statistical weights ; similar

conditions would presumably apply in other fine structure states. Because it still seems uncertain, however, whether cross sections computed in the above way yield reliable results for states of very small energy difference, we have chosen to neglect $2S \rightarrow 2P$ transitions. From results obtained here, it would appear that this type of transition will have little effect, in any case, except at the higher temperatures where the $2P$ and $2S$ populations are markedly different.

TABLE 3

RATES OF SPONTANEOUS RECOMBINATION

$(10^{-15}N_e{}^2 \text{ cm.}^{-3} \text{ sec.}^{-1})$

T (°K.)	$i \rightarrow 1S$	$i \rightarrow 2S$	$i \rightarrow 3S$
$1 \cdot 0 \times 10^4$	$2 \cdot 07 \times 10^2$	$2 \cdot 15 \times 10$	$5 \cdot 20$
$1 \cdot 5 \times 10^4$	$1 \cdot 56 \times 10^2$	$1 \cdot 60 \times 10$	$3 \cdot 77$
$2 \cdot 5 \times 10^4$	$1 \cdot 14 \times 10^2$	$1 \cdot 11 \times 10$	$2 \cdot 52$
$5 \cdot 0 \times 10^4$	$7 \cdot 30 \times 10$	$6 \cdot 36$	$1 \cdot 31$
$1 \cdot 0 \times 10^5$	$4 \cdot 51 \times 10$	$3 \cdot 41$	$6 \cdot 68 \times 10^{-1}$
$2 \cdot 5 \times 10^5$	$2 \cdot 10 \times 10$	$1 \cdot 36$	$2 \cdot 44 \times 10^{-1}$

However, the emergent Hα intensity is unlikely to be significantly affected, even at these temperatures, because of the small optical depth ; the Lα intensity computed here may be slightly low.

V. THE POPULATION OF THE BASE STATE

In a model chromosphere at 25,000 °K., the ratios of the $2P$ and $3P$ populations to that of the ground state are effectively maintained by the absorption and re-emission of Lyman line radiation (Giovanelli 1949) ; i.e. the rates of radiative excitation $\Lambda_\alpha N_1$ and $\Lambda_\beta N_1$ defined by

$$\Lambda_\alpha N_1 = \int J_\nu \alpha_\nu d\nu / h\nu \text{ and } \Lambda_\beta N_1 = \int J_\nu \alpha_\nu d\nu / h\nu$$

may be respectively equated to $A_{2P,1S}N_{2P}$ and $A_{3P,1S}N_{3P}$, where $A_{j,l}$ is the appropriate spontaneous transition probability. The same results apply to the cases considered here, provided the atmosphere is optically thick to the Lyman line radiations. In this case the total intensities of Lα and Lβ inside the chromosphere will be given by $J = 4\pi\varepsilon/\alpha\lambda$.

We now evaluate N_1 on the assumption that this relation holds (in the chromosphere) for all electron temperatures ; this assumption will be shown to break down for the higher temperatures, for which a different procedure is required.

For any given temperature, it may then be shown (Giovanelli 1949) that

$$N_1 = \frac{AN_e{}^2}{\Lambda_c + BN_e} \quad \ldots\ldots\ldots\ldots\ldots\ldots (3)$$

to within a factor less than 2 over the range of N_e from 10^{10} to 10^{12} cm.$^{-3}$, $\Lambda_c N_1$ being the rate of photoelectric ionization from the ground state per unit volume,

and A and B numerical constants obtained directly from the equilibrium equations. Value of BN_e are shown in Table 6. If Λ_c is neglected in comparison with BN_e, a procedure whose validity is discussed later, the approximate values of N_1 shown in Table 4 are obtained.

TABLE 4

BASE STATE POPULATIONS : FIRST APPROXIMATION
(SEE TEXT)

T (°K.)	N_1 (cm.$^{-3}$)
$1 \cdot 0 \times 10^4$	$6 \cdot 86 \times 10^{-1} N_e$
$1 \cdot 5 \times 10^4$	$8 \cdot 78 \times 10^{-3} N_e$
$2 \cdot 5 \times 10^4$	$2 \cdot 09 \times 10^{-4} N_e$
$5 \cdot 0 \times 10^4$	$8 \cdot 78 \times 10^{-6} N_e$
$1 \cdot 0 \times 10^5$	$1 \cdot 28 \times 10^{-6} N_e$
$2 \cdot 5 \times 10^5$	$2 \cdot 63 \times 10^{-7} N_e$

The total optical depth of the chromosphere at the centres of the Lyman lines and at the Lyman series limit may readily be calculated, using the approximate base state populations given in Table 4, from the relation

$$\tau_1{}^0 = -\int_0^\infty \alpha_{\nu_0} dz, \quad \dots\dots\dots\dots\dots \quad (4)$$

where

$$\alpha_{\nu_0} = 2 \cdot 83 \times 10^{29} N_1/\nu_0{}^3 \quad \text{per cm. for Lyman series limit,}$$
$$= 6 \cdot 02 \times 10^{29} N_1/\sqrt{T} \quad \text{per cm. for the centre of L}\alpha,$$

and

$$= 9 \cdot 66 \times 10^{-13} N_1/\sqrt{T} \quad \text{per cm. for the centre of L}\beta.$$

Values of the optical depth in the Lyman continuum and of the quantities $\sqrt{3\lambda}\tau_1{}^\circ$ for Lα and Lβ are shown in Table 5 for $N_0 = 5 \times 10^{11}$ cm.$^{-3}$, $\beta = 6 \times 10^{-9}$ cm.$^{-1}$, N_0 being the electron concentration at the base of the chromosphere, and β the electron gradient.

TABLE 5

OPTICAL DEPTH OF HYPOTHETICAL CHROMOSPHERES FOR LYMAN RADIATION (SEE TEXT)

T (°K.)	$\sqrt{3\lambda}\tau_1{}^\circ$ (Lα)	$\sqrt{3\lambda}\tau_1{}^\circ$ (Lβ)	$\tau_1{}^\circ$ (Lyman Limit)
$1 \cdot 0 \times 10^4$	$5 \cdot 85 \times 10^4$	$9 \cdot 90 \times 10^4$	$4 \cdot 53 \times 10^2$
$1 \cdot 5 \times 10^4$	$7 \cdot 30 \times 10^2$	$1 \cdot 23 \times 10^3$	$5 \cdot 82$
$2 \cdot 5 \times 10^4$	$1 \cdot 84 \times 10$	$2 \cdot 76 \times 10$	$1 \cdot 40 \times 10^{-1}$
$5 \cdot 0 \times 10^4$	$6 \cdot 65 \times 10^{-1}$	$9 \cdot 70 \times 10^{-1}$	$5 \cdot 78 \times 10^{-3}$
$1 \cdot 0 \times 10^5$	$7 \cdot 32 \times 10^{-2}$	$8 \cdot 51 \times 10^{-2}$	$8 \cdot 45 \times 10^{-4}$
$2 \cdot 5 \times 10^5$	$9 \cdot 38 \times 10^{-3}$	$1 \cdot 18 \times 10^{-2}$	$1 \cdot 73 \times 10^{-4}$

It can be seen that the optical depth of the chromosphere is appreciable in the Lyman continuum, for temperatures of $1 \cdot 5 \times 10^4$ °K. or less, and may be small for the Lyman lines when the temperature is 5×10^4 °K. or more, depending on the value of N_1. We shall now reconsider the values of N_1 in such cases.

(a) Effect of the Lyman Continuum

Neglecting reflection at the base of the chromosphere, the total monochromatic intensity in the Lyman continuum, J_ν, at the top of the chromosphere is given by the relation (see Giovanelli 1949, equation (73)):

$$J_\nu = \pi G \beta H^{-2}[1 - (1+a) \exp(-a) - a^2 \mathrm{Ei}(-a)], \quad \dots \text{(5)}$$

where

$$G = 1 \cdot 72 \times 10^{-33} T^{-3/2} \exp[(E_1^i - h\nu)/kT],$$
$$H = 2 \cdot 83 \times 10^{29} N_1/N_e \nu^3,$$
$$a = N_o H \beta^{-1},$$

and

$$-\mathrm{Ei}(-a) = \int_a^\infty [\exp(-x)/x]\mathrm{d}x.$$

E_1^i is the ionization potential and the other symbols have their usual meanings.

If $T \geqslant 2 \cdot 5 \times 10^4$ °K. it may be seen from Table 5 that self absorption in the Lyman continuum will be small and so the intensity of this radiation inside the chromosphere may be taken equal to that escaping from the top, which is given by equation (5).

When T is appreciably less than $2 \cdot 5 \times 10^4$ °K., the optical depth becomes much greater than unity and so the intensity J_ν will be approximately equal to $4\pi B_\nu$, where B_ν is defined as the ratio of the total emission coefficient to the absorption coefficient and for the Lyman continuum is given by (cf. Giovanelli 1949, equation (20))

$$B_\nu = \frac{6 \cdot 1 \times 10^{-63} N_e^2 \nu^3 \exp(E_1^i - h\nu)/kT}{N_1 T^{3/2}}. \quad \dots\dots\dots\dots \text{(6)}$$

Assuming for simplicity that the function $f(a) = 1 - (1+a)e^{-a} - a^2\mathrm{Ei}(-a)$ of equation (5) is independent of frequency, which introduces no great error, and substituting numerical values we find that

$$\Lambda_c = \frac{1}{N_1} \int_{\nu_0}^\infty \frac{J_\nu \alpha_\nu \mathrm{d}\nu}{h\nu} = 1 \cdot 19 \times 10^9 J_0[b + 2b^2 + b^3], \quad \dots\dots\dots \text{(7)}$$

for $T \geqslant 2 \cdot 5 \times 10^4$ °K., where $b = kT/h\nu_0$ and J_0 is the intensity at the head of the Lyman continuum.

For $T < 2 \cdot 5 \times 10^4$ °K., $J_\nu \sim 4\pi B_\nu$ inside the chromosphere, so that

$$\Lambda_c = \frac{3 \cdot 25 \times 10^{-7} N_e^2}{N_1 T^{3/2}} \exp(E_1^i/kT)\{-\mathrm{Ei}(-E_1^i/kT)\}.$$

Thus, since $E_1^i/kT \gg 1$ at these temperatures

$$\Lambda_c \simeq \frac{3 \cdot 25 \times 10^{-7} N_e^2}{N_1 T^{3/2}} \cdot \frac{kT}{E_1^i}. \quad \dots\dots\dots\dots \text{(8)}$$

Values of Λ_c obtained in this way are given in Table 6 for two values of the electron concentration. For $T \geqslant 2\cdot5\times10^4$ °K., where Λ_c is in general much less than BN_e, photoelectric ionization from the ground state may clearly be neglected.

At the lower temperatures, 1×10^4 and $1\cdot5\times10^4$ °K., values of Λ_c obtained by using the approximate N_1 given in Table 4 are of the same order as the BN_e terms. We may find both Λ_c and N_1 in this case by combining equations (3) and (8). Ground state populations determined in this way are as follows :

T (°K.)	N_1
1×10^4	$6\cdot09\times10^{-1}N_e$
$1\cdot5\times10^4$	$7\cdot33\times10^{-3}N_e$

These values are little different from those of Table 4.

TABLE 6

COMPARATIVE VALUES OF Λ_c AND BN_e (SEE TEXT)

T (°K.)	Λ_c		BN_e
	$N_e = 5\times10^{11}$	$N_e = 10^{12}$	
$1\cdot0\times10^4$	$0\cdot0171$	$0\cdot0342$	$3\cdot63\times10^{-13}N_e$
$1\cdot5\times10^4$	$2\cdot12$	$4\cdot24$	$2\cdot13\times10^{-11}N_e$
$2\cdot5\times10^4$	$16\cdot4$	64	$6\cdot52\times10^{-10}N_e$
$5\cdot0\times10^4$	42	149	$9\cdot65\times10^{-9}N_e$
$1\cdot0\times10^5$	110	364	$3\cdot96\times10^{-8}N_e$
$2\cdot5\times10^5$	92	347	$8\cdot67\times10^{-8}N_e$

(b) Effect of the Lyman Line Radiations

For $T \geqslant 5\times10^4$ °K. the ground state populations will be greater than those shown in Table 4, since the rate at which atoms are removed from this state by absorption of Lα and Lβ radiation is less than the rate of entry from the $3P$ and $2P$ states. We shall now compute N_1 in this case.

The intensity of Lyman line radiation at the base of the chromosphere may be seen from results of Giovanelli (1949) to be given approximately by :

$$J_\nu = \frac{4\pi\varepsilon}{\alpha\lambda}\left[1 - \frac{2\exp\left(-2\tau_1\sqrt{3\lambda}\right)}{1+2\sqrt{\lambda/3}+(1-2\sqrt{\lambda/3})\exp\left(-2\tau_1\sqrt{3\lambda}\right)}\right], \quad ..(9)$$

which may be written as

$$J_\nu = k(4\pi\varepsilon/\alpha\lambda).$$

The values of $\Lambda\alpha$ and $\Lambda\beta$ to be used in calculating the constants A and B of equation (3) should then be no greater than k times the values obtained by assuming the atmosphere to be optically thick to Lα and Lβ.

Using Table 5, the approximate values of the k's may be calculated, and are found to be given, for the line centres, by $k\alpha \simeq k\beta = 10^{-2}$ for $T=10^5$ °K. and $k\alpha \simeq k\beta = 10^{-4}$ for $T=2\cdot5\times10^5$ °K.

Substituting the new values of Λ_α, Λ_β in the equations for N_1 we find

T (°K.)	N_1
1×10^5	$3 \cdot 50 \times 10^{-6} N_e$
$2 \cdot 5 \times 10^5$	$7 \cdot 38 \times 10^{-7} N_e$

Although these values of N_1 will, in turn, modify the values of k, the resulting change in N_1 is insignificant.

TABLE 7

BASE   STATE   POPULATIONS   FOR   HYPOTHETICAL
CHROMOSPHERES (SEE TEXT)

T (°K.)	N_1
$1 \cdot 0 \times 10^4$	$6 \cdot 09 \times 10^{-1} N_e$
$1 \cdot 5 \times 10^4$	$7 \cdot 33 \times 10^{-3} N_e$
$2 \cdot 5 \times 10^4$	$2 \cdot 09 \times 10^{-4} N_e$
$5 \cdot 0 \times 10^4$	$8 \cdot 78 \times 10^{-6} N_e$
	(or $2 \cdot 0 \times 10^{-5} N_e$)
$1 \cdot 0 \times 10^5$	$3 \cdot 50 \times 10^{-6} N_e$
$2 \cdot 5 \times 10^5$	$7 \cdot 38 \times 10^{-7} N_e$

The values of N_1 adopted here for $N_e \leqslant 10^{12}$ cm.$^{-3}$ are set out in Table 7. These values include the effect of photoelectric ionization at low temperatures and of the small optical depth in the Lyman lines at high temperatures.   The results should be correct to within a factor of about two.

For $T = 5 \times 10^4$ °K. (where the optical depths in Lα and Lβ are about unity), the value of N_1 shown in brackets has been used for $N_e \leqslant 5 \times 10^{11}$ cm.$^{-3}$; this is the value obtained by application of the appropriate correction factors to Λ_α and Λ_β, as above.

VI. THE LYMAN CONTINUUM

The values of N_1 shown in Table 7 may be used, with specified values of N_0, β, and temperature, to calculate J_ν from equation (5).   These results, expressed in terms of the equivalent temperature for hemispherical radiation from a black body, are shown graphically in Figure 1.   Giovanelli's value for $T = 2 \cdot 5 \times 10^4$ °K. and $N_0 = 10^{12}$ cm.$^{-3}$ is also plotted and it will be seen that the improved cross sections have not significantly affected the results.

Emission of Lyman continuous radiation reaches a maximum at a temperature of about 5×10^4 °K.   At high temperatures, where the atmosphere is effectively transparent, the emission depends only on recombination and so falls off as $T^{-\frac{1}{2}}$.

VII. THE LYMAN LINES

The intensity of the escaping radiation in the Lyman lines is given (Giovanelli 1949) by

$$J_\nu = \frac{4\pi\varepsilon}{\alpha\lambda} \left[1 - \frac{1 + \exp\left(-2\tau_1\sqrt{3\lambda}\right)}{(1 + 2\sqrt{\lambda/3}) + (1 - 2\sqrt{\lambda/3})\exp\left(-2\tau_1\sqrt{3\lambda}\right)} \right].$$

$$\cdots\cdots\cdots\cdots (10)$$

On substituting numerical values of ε/α and λ (estimated by the method used by Giovanelli (1949)) into (**10**), the central intensities of the Lyman lines shown in Figure 2 are obtained for $N_0 = 5 \times 10^{11}$ cm.$^{-3}$, $\beta = 6 \times 10^{-9}$ cm.$^{-1}$. As
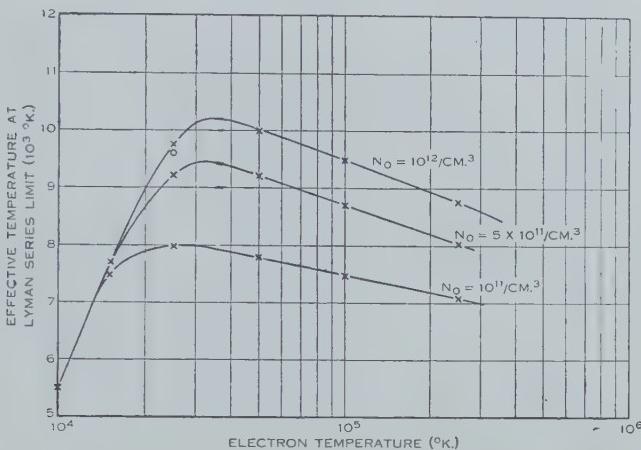
Fig. 1.—Effective temperature at the beginning of the Lyman continuum for $N_0 = 10^{11}$, 5×10^{11}, and 10^{12} cm.$^{-3}$, $\beta = 6 \times 10^{-9}$ cm.$^{-1}$.

in the case of the Lyman continuum, the intensity of the emergent radiation passes through a maximum at $T \simeq 5 \times 10^4$ °K.

The contour of a Lyman line may also be obtained from equation (**10**). Calculations show that λ does not vary very rapidly with electron temperature,
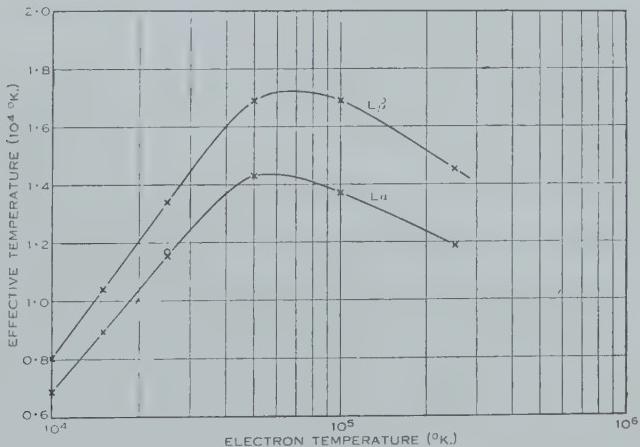
Fig. 2.—Effective temperature at the centre of Lα and Lβ. $N_0 = 5 \times 10^{11}$ cm.$^{-3}$, $\beta = 6 \times 10^{-9}$ cm.$^{-1}$.

and so the factor governing the intensity at a frequency of $\nu_0 \pm \Delta\nu$ is the optical depth, which takes the form $PN_0 \exp[-\gamma(\Delta\nu)^2]$, where P and γ are quantities which decrease with increasing T.

c

At the lower temperatures for which $\sqrt{3\lambda}\,\tau_1{}^\circ \gg 1$, the ratio $J_0/(4\pi\varepsilon/\alpha\lambda)$ is effectively independent of T, and so the frequency at which this ratio is reduced to one-half will be such that $\sqrt{3\lambda}\,\tau_1$ is almost constant with T.

For high electron temperatures where $\sqrt{3\lambda}\,\tau_1{}^\circ \ll 1$, the profile of the emission line is undisturbed by self absorption, and is, at least not too far out in the wings, of the Doppler form. Consequently, at half-width, $\exp[-\gamma(\Delta\nu)^2] = \frac{1}{2}$, and since γ varies as T^{-1}, the half-width of the line varies as $T^{\frac{1}{2}}$. The term half-width is used for Lα and Lβ to denote the distance from the line centre at which the intensity has dropped to one-half of its value at the centre (our half-width is

TABLE 8

HALF-WIDTHS OF Lα AND Lβ

$N_0 = 5 \times 10^{11}$ cm.$^{-3}$, $\beta = 6 \times 10^{-9}$ cm.$^{-1}$

T (°K.)	Lα (Å)	Lβ (Å)
$1 \cdot 0 \times 10^4$	$0 \cdot 18$	$0 \cdot 15_5$
$1 \cdot 5 \times 10^4$	$0 \cdot 17$	$0 \cdot 15$
$2 \cdot 5 \times 10^4$	$0 \cdot 16$	$0 \cdot 14$
$5 \cdot 0 \times 10^4$	$0 \cdot 12$	$0 \cdot 11$
$1 \cdot 0 \times 10^5$	$0 \cdot 14$	$0 \cdot 12$
$2 \cdot 5 \times 10^5$	$0 \cdot 22$	$0 \cdot 18$

sometimes known as the " half-half-width "). The half-width increases with the value of the electron concentration at the base of the chromosphere— through the neutral atom concentration—for all temperatures, although the increase with N_0 may become negligible at high temperatures. The calculated half-widths, in Ångstrom units, for the Lα and Lβ lines are tabulated in Table 8 for $N_0 = 5 \times 10^{11}$ cm.$^{-3}$, $\beta = 6 \times 10^{-9}$ cm.$^{-1}$.

VIII. The Hα Intensity

Whereas the Lyman lines are of much greater intensity than the neighbouring continuous spectrum from the photosphere, the Hα line appears dark against a brighter photospheric continuous spectrum. The computation of the Hα contour is thus much more complicated, in that photospheric radiation cannot be ignored. The contour, which depends on the optical depth of the chromosphere, may be determined either by the chromosphere, by the photosphere, or, as is more probable, by a combination of both. In general, the chromosphere exerts its maximum effect near the centre of the line, where its optical depth is greatest. In the far wings, where the chromosphere is transparent, the computation of the contour will need to take account of the change with wavelength of the photospheric absorption and scattering coefficients.

In investigating these matters, we do not propose to consider in detail the processes of photospheric absorption and scattering, nor to take into account

variations in physical conditions with optical depth. It is sufficient to assume, in the photosphere, B_v to be constant, and that the scattering parameter λ_v' does not change with optical depth. In the chromosphere we again take $4\pi\varepsilon/\alpha\lambda$ and λ to be constant both with frequency and with optical depth. The approximate solution of the equation of radiative transfer then becomes

$$J_c = \frac{4\pi\varepsilon}{\alpha\lambda} + a\exp\left(\sqrt{3\lambda}\tau\right) + b\exp\left(-\sqrt{3\lambda}\tau\right). \quad \ldots\ldots (11)$$

in the chromosphere, and

$$J_p = 4\pi B_v + \gamma\exp\left(-\sqrt{3\lambda'}\tau'\right) \quad \ldots\ldots\ldots\ldots\ldots (12)$$

in the photosphere, a, b, and γ being integration constants and τ and τ' being optical depths measured from the top of the chromosphere and photosphere respectively. For the sake of clarity, frequency subscripts have been omitted.

To obtain the required intensity of the radiation in the chromosphere we make use of the boundary conditions that (i) across the interface of the two atmospheres the intensities and fluxes are continuous, and (ii) at the outer boundary of the chromosphere there is no incident radiation. These conditions are equivalent respectively to

$$J_c = J_p, \quad \frac{\partial J_c}{\partial\tau} = \frac{\partial J_p}{\partial\tau}.$$

at the interface, and

$$\frac{2}{3}\frac{\partial J_c}{\partial\tau} = J_c \quad \text{at} \quad \tau = 0.$$

With these conditions we find that

$$a = -\frac{4\pi\varepsilon/\alpha\lambda + (1 + 2\sqrt{\lambda/3})b}{1 - 2\sqrt{\lambda/3}}, \quad \ldots\ldots\ldots\ldots\ldots\ldots (13)$$

$$a = -\frac{\sqrt{\lambda'}(4\pi\varepsilon/\alpha\lambda - 4\pi B_0) + b(\sqrt{\lambda'} - \sqrt{\lambda})\exp\left(-\sqrt{3\lambda}\tau_1\right)}{(\sqrt{\lambda} + \sqrt{\lambda'})\exp\left(\sqrt{3\lambda}\tau_1\right)},$$

$$\ldots\ldots\ldots\ldots\ldots\ldots (14)$$

where τ_1 is the total optical depth of the chromosphere and B_0 is the value of B_v at the top of the photosphere.

At the outer boundary $J_c = 4\pi\varepsilon/\alpha\lambda + a + b$. Substituting values of a and b found from the solutions of the equations (13) and (14), we find that the emergent intensity is given by

$$J_c = J_0 + J_1, \quad \ldots\ldots\ldots\ldots\ldots\ldots (15)$$

where

$$J_0 = \frac{4\pi\varepsilon}{\alpha\lambda}\left[1 - \frac{\sqrt{\lambda} + \sqrt{\lambda'} + (\sqrt{\lambda} - \sqrt{\lambda'})\exp\left(-2\tau_1\sqrt{3\lambda}\right)}{(1 + 2\sqrt{\lambda/3})(\sqrt{\lambda} + \sqrt{\lambda'}) + (1 - 2\sqrt{\lambda/3})(\sqrt{\lambda} - \sqrt{\lambda'})\exp\left(-2\tau_1\sqrt{3\lambda}\right)}\right]$$

and

$$J_1 = -\frac{4\sqrt{\lambda/3}\sqrt{\lambda'}\exp\left(-\sqrt{3\lambda}\tau_1\right)[4\pi\varepsilon/\alpha\lambda - 4\pi B_0]}{(1 + 2\sqrt{\lambda/3})(\sqrt{\lambda} + \sqrt{\lambda'}) + (1 - 2\sqrt{\lambda/3})(\sqrt{\lambda} - \sqrt{\lambda'})\exp\left(-2\tau_1\sqrt{3\lambda}\right)}.$$

In estimating the value of B_0 in (15), we may note that over the wavelength range concerned, B_0 is independent of frequency. In the far wings of Hα, where τ tends to zero,

$$J_c = \frac{2\sqrt{\lambda'/3}}{1+2\sqrt{\lambda'/3}} \times 4\pi B_0.$$

Since, as will be shown later, $\lambda'=1$ here,

$$J_c = \frac{4\pi B_0}{1+\sqrt{3}/2}. \quad \dots\dots\dots\dots\dots\dots (16)$$

This, however, represents the emergent continuous radiation which for the case of the Sun at Hα approximates to the hemispherical radiation from a black body at 6150 °K. Denoting this by J_w, we have, from (16),

$$4\pi B_0 = (1+\sqrt{3}/2)J_w. \quad \dots\dots\dots\dots\dots\dots (17)$$

Finally, substituting for $4\pi B_0$ from this equation into equation (15) we find

$$J_c =$$
$$\frac{4\pi\varepsilon}{\alpha\lambda}\left[\frac{1-(\sqrt{\lambda}+\sqrt{\lambda'})+4\sqrt{\lambda\lambda'/3}\exp(-\sqrt{3}\lambda\tau_1)+(\sqrt{\lambda}-\sqrt{\lambda'})\exp(-2\tau_1\sqrt{3\lambda})}{D}\right]$$
$$+\frac{2(1+2/\sqrt{3})\sqrt{\lambda\lambda'}\exp(-\sqrt{3\lambda}\tau_1)J_w}{D}, \quad \dots\dots\dots\dots\dots\dots (18)$$

where

$$D = (1+2\sqrt{\lambda/3})(\sqrt{\lambda}+\sqrt{\lambda'})+(1-2\sqrt{\lambda/3})(\sqrt{\lambda}-\sqrt{\lambda'})\exp(-2\tau_1\sqrt{3\lambda}).$$

We shall now consider the evaluation of ε/α and of λ and λ' in equation (18).

(a) The Ratio ε/α

The ratio ε/α depends to some extent on the Hα intensity, a lower intensity allowing a higher population in the metastable $2S$ state. Consequently the evaluation of ε/α and α from the equilibrium equation must be in terms not only of atomic constants, but also of the rate of absorption of Hα quanta by atoms in the $2S$ state, which may be written

$$\Lambda_{2S-3P}N_{2S} = \int\frac{J_\nu\alpha_\nu d\nu}{h\nu}. \quad \dots\dots\dots\dots\dots (19)$$

Since the maximum value of α_ν occurs at the centre of the line, and J_ν varies less rapidly with ν than does α_ν, the integral is given quite closely by $(J_0/h\nu_0)\int\alpha_\nu d\nu$ where J_0 is the intensity at the centre of the line.

The emergent radiation at the centre of the line comes, on the average, from regions where $\sqrt{3}\lambda\tau° = 1$, and a study of conditions at this level enables ε/α, and consequently the Hα intensity, to be estimated. The approximate electron concentration at this level has been obtained by assuming a suitable value for the Hα intensity—in the case of the Sun this corresponds approximately to that inside a black body at 5000 °K.—calculating $\sqrt{3}\lambda\tau°$ in terms of N_e, and finding N_e for $\sqrt{3}\lambda\tau°=1$. For $\sqrt{3}\lambda\tau_1°<1$ the values of ε/α and α have been computed for the electron concentration at the base of the chromosphere.

(b) The Scattering Parameters

The photospheric scattering parameter λ' is defined by the equation

$$1-\lambda\nu'=\frac{\sigma\nu}{\sigma\nu+\varkappa\nu+\varkappa_0}, \quad \dots\dots\dots\dots\dots \quad (20)$$

where $\sigma\nu$ and $\varkappa\nu$ are the coefficients of scattering and " true " absorption for the Hα line, and $\varkappa_0$ is the coefficient of continuous absorption by negative hydrogen ions, at the frequency ν. Values of $\varkappa_0$ may be obtained from tables given by Chandrasekhar and Breen (1946). To evaluate $\lambda\nu'$ we make use of λ_L, the line scattering parameter defined by the equation

$$1-\lambda_L=\frac{\sigma\nu}{\sigma\nu+\varkappa\nu}. \quad \dots\dots\dots\dots\dots \quad (21)$$

Then from equation (20)

$$\lambda\nu'=\frac{\sigma\nu+\varkappa\nu}{\sigma\nu+\varkappa\nu+\varkappa_0}\lambda_L+\frac{\varkappa_0}{\sigma\nu+\varkappa\nu+\varkappa_0}$$

$$=\frac{\alpha\nu\lambda_L}{\alpha\nu+\varkappa_0}+\frac{\varkappa_0}{\alpha\nu+\varkappa_0}, \quad \dots\dots\dots\dots\dots \quad (22)$$

as $\sigma\nu+\varkappa\nu=\alpha\nu$. We see at once that, as $\alpha\nu\to0$, $\lambda\nu'\to1$.

To estimate λ_L we make use of the result obtained by calculation that the Hα scattering parameter in the chromosphere has a value (about $0\cdot35$) which is relatively insensitive to variations in N_e or T : for the high rate of spontaneous $3P\to1S$ transition, which is independent of the physical conditions, determines its magnitude. For higher N_e, collision transitions from the 3-quantum state would become increasingly more frequent and λ approach unity. Thus λ_L will differ from λ only if collision transitions from the 3 states are much more important in the photosphere than in the chromosphere.

Using cross-section data based on general theorems due to Bohr, Peierls, and Placzek (1949) for low energy inelastic collisions, it may be readily shown that, in the photosphere, transitions from the $3P$ state are still predominantly spontaneous provided the electron concentration does not much exceed 10^{13} cm.$^{-3}$. In the layers responsible for the radiation in the wings of Hα, however, N_e may exceed this value, and it is a little difficult to see what λ_L will be. It is possible, however, to obtain its value in the wings from observation, as shown later, and as this also turns out to be about $0\cdot35$, it presumably applies closer to the line centre, where collision transitions will be of less importance.

In the wings of Hα the optical depth of the chromosphere is effectively zero, and the emergent intensity can be obtained by substituting $\tau_1=0$ in equation (18), when we find

$$\frac{J\nu}{J_w}=1\cdot866\frac{2\sqrt{\lambda\nu'/3}}{1+2\sqrt{\lambda\nu'/3}}. \quad \dots\dots\dots\dots\dots \quad (23)$$

Hence

$$\lambda\nu'=\frac{3}{4}\left[\frac{r\nu}{1\cdot866-r\nu}\right]^2, \quad \dots\dots\dots\dots\dots \quad (24)$$

where $r\nu=J\nu/J_w$.

Far enough away from the line centre, the Hα absorption coefficient may be written

$$\alpha_\nu = b/(\Delta\nu)^2, \quad \dots\dots\dots\dots\dots\dots\dots \text{(25)}$$

where b is a function of N_{2S}, N_{2P}, and N_e. Its dependence on N_e arises from the Stark broadening of the energy levels, compared with which natural broadening may generally be neglected in the photosphere.

From (25) and (22) we find

$$\lambda_\nu' = \frac{\lambda_L + gx}{1 + gx},$$

or

$$\frac{1}{1-\lambda_\nu'} = \frac{gx}{1-\lambda_L} + \frac{1}{1-\lambda_L}, \quad \dots\dots\dots\dots \text{(26)}$$

where $g = \varkappa_0/b$ and $x = (\Delta\nu)^2$.

This relation may now be compared with observations of the contour of Hα emitted by the sun.
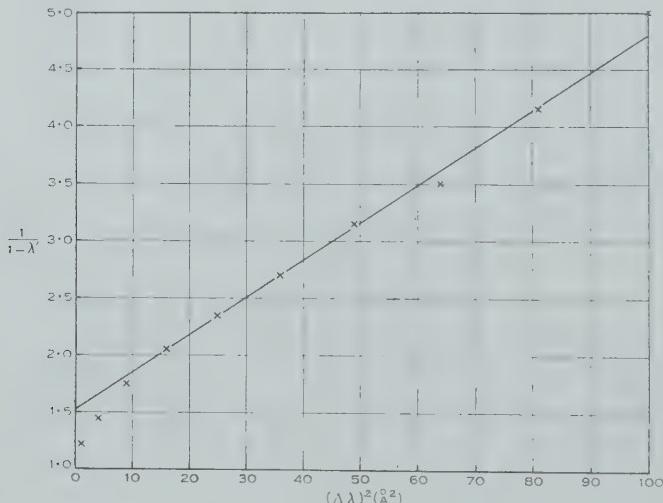
Fig. 3.—Graph of $1/(1-\lambda')$ v. $(\Delta\lambda)^2$ for Hα ; see text.

Figure 3 shows $1/(1-\lambda_\nu')$, obtained from Evans's (1940) results via equation (24), plotted against $(\Delta\nu)^2$. From the linear nature of the graph it would appear that, beyond 4 Å from the centre of the line, the effective values of λ_L and $\varkappa_0/b$ are constant, being $0 \cdot 35$ and $4 \cdot 3 \times 10^{-24}$ sec.^{2}. Using data given by Chandra-sekhar and Breen (1946) and well-known expressions for Stark broadening (see, e.g. White 1934), the ratio $\varkappa_0/b$ may be calculated, for a gas in thermodynamic equilibrium, in terms of N_e ($=N_+$) and T. The value found above is numerically equal to that which would be obtained for an atmosphere in thermodynamic equilibrium at about 5700 °K. if the energy levels are Stark broadened and $N_e = 10^{12}$ cm.$^{-3}$, and 6000 °K. if $N_e = 10^{13}$ cm.$^{-3}$.

Equation (26) may be used to calculate λ_ν' from λ_L and g. It is clear that if the wing intensities are now computed from equation (23), they must fit

Evans's experimental results over the region of wavelengths where the curve of Figure 3 is linear.

Evans's results from the Hα contour were used in plotting Figure 3, since they provide data at a large number of wavelengths. The photospheric parameters have been also computed from the means of four sets of observations
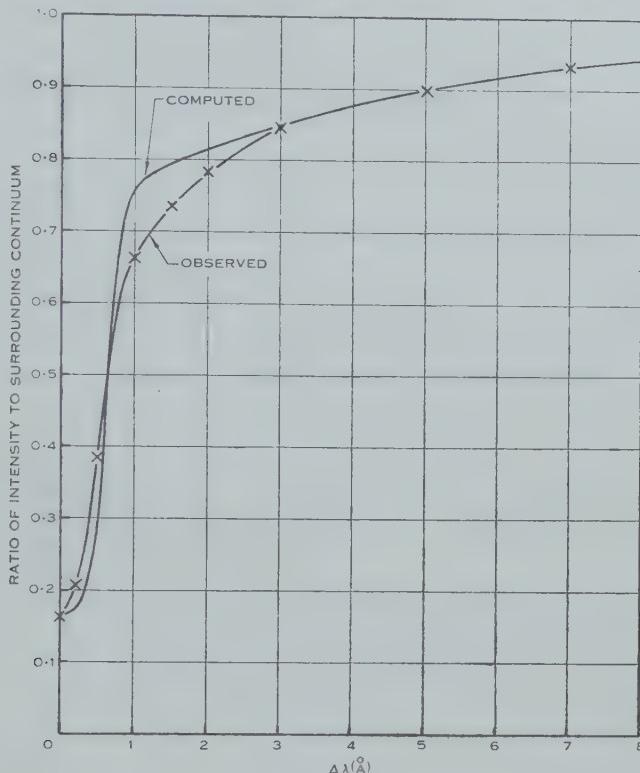
Fig. 4.—Hα contour calculated for $T = 2.5 \times 10^4$ °K., $N_0 = 2 \times 10^{11}$ cm.$^{-3}$, $\beta = 6 \times 10^{-9}$ cm.$^{-1}$, compared with observational results.

(Minnaert 1927 ; Thackeray 1935 ; Evans 1940 ; ten Bruggencate et al. 1949), the resultant values of α_ν and λ_L obtained being

$$\alpha_\nu = \frac{1 \cdot 62 \times 10^{14}}{(\Delta\nu)^2},$$

$$\lambda_L = 0 \cdot 37,$$

for

$$\varkappa_0 = 10^{-9} \text{ cm.}^{-1}.$$

This value of λ_L is closely equal to the value computed for the chromosphere, so that λ_L may be taken to be constant for the whole line.

Figures 4 and 5 show Hα contours computed from these data for $T = 2 \cdot 5 \times 10^4$ °K., $N_0 = 2 \times 10^{11}$ cm.$^{-3}$, and $\beta = 6 \times 10^{-9}$ cm.$^{-1}$. Figure 5 also shows the effect of an increased electron concentration at the base of the chromosphere. The shape of the line is clearly sensitive to N_0.

Central intensities of the emergent Hα line have been computed over the full range of temperatures, with the approximation $\lambda_{\nu_0}' = \lambda$. For a given electron concentration N_0 at the base of the chromosphere, the intensity depends,
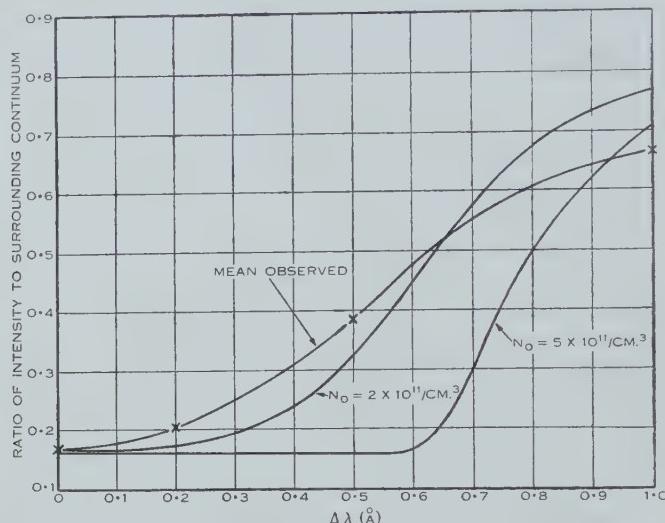
Fig. 5.—Contour of Hα near the centre of the line calculated for $N_0 = 2 \times 10^{11}$ cm.$^{-3}$ and $N_0 = 5 \times 10^{11}$ cm.$^{-3}$ with $T = 2.5 \times 10^4$ °K., $\beta = 6 \times 10^{-9}$ cm.$^{-1}$.

through the optical depth, on the electron gradient β. Results of calculations are shown in Table 9 for $\beta = 6 \times 10^{-9}$ cm.$^{-1}$ and in Table 10 for an atmosphere in hydrostatic equilibrium. These tables show the contributions of the chromospheric and photospheric components to the central intensity of the emergent

TABLE 9

CONTRIBUTIONS OF CHROMOSPHERE AND PHOTOSPHERE TO Hα CENTRAL INTENSITY

$\beta = 6 \times 10^{-9}$ cm.$^{-1}$

T (°K.) \ N_0 (cm.$^{-3}$)	5×10^{12}		10^{12}		5×10^{11}		10^{11}	
1.0×10^4	9	0	9	0	9	0	9	0
1.5×10^4	12	0	12	0	12	0	12	2
2.5×10^4	16	0	16	0	16	0	14	16
5.0×10^4	62	0	62	0	42	25	2	68
1.0×10^5	128	0	50	43	14	59	0	73
2.5×10^5	81	63	0	76	0	76	0	76

Hα radiation in units such that the intensity of the surrounding continuum equals 100. The chromospheric component is shown on the left of each column. The total central intensity is given by the sum of the two components.

The central intensities depend upon the optical depths and for the lower temperatures, $T \leqslant 2 \cdot 5 \times 10^4$ °K., should be fairly reliable.   For $T > 2 \cdot 5 \times 10^4$ °K., however, the optical depth is more difficult to estimate owing to the low value

TABLE 10

CONTRIBUTIONS OF CHROMOSPHERE AND PHOTOSPHERE TO Hα CENTRAL INTENSITY

HYDROSTATIC EQUILIBRIUM

T (°K.)	N_0 (cm.$^{-3}$) 10^{12}		5×10^{11}		10^{11}	
$1 \cdot 0 \times 10^4$	9	0	9	0	9	0
$1 \cdot 5 \times 10^4$	12	0	12	0	12	2
$2 \cdot 5 \times 10^4$	16	0	16	0	14	16
$5 \cdot 0 \times 10^4$	30	0	30	0	9	61
$1 \cdot 0 \times 10^5$	68	0	41	28	0	73
$2 \cdot 5 \times 10^5$	30	62	2	73	0	76

of $\sqrt{3\lambda}\tau_1°$ in the Lyman lines, making the calculated ratios N_{3P}/N_{1S} and N_{2P}/N_{1S} somewhat uncertain.   Nevertheless the errors in central intensity should not be very great, as the errors in photospheric and chromospheric contributions to some extent balance one another.

TABLE 11

CENTRAL INTENSITIES AND HALF-WIDTHS OF Hα AT VARIOUS CHROMOSPHERIC

TEMPERATURES

$N_0 = 5 \times 10^{11}$ cm.$^{-3}$, $\beta = 6 \times 10^{-9}$ cm.$^{-1}$

T (°K.)	Central Intensity	Half-widths (Å)
$1 \cdot 0 \times 10^4$	$0 \cdot 09$	$0 \cdot 64$
$1 \cdot 5 \times 10^4$	$0 \cdot 12$	$0 \cdot 71$
$2 \cdot 5 \times 10^4$	$0 \cdot 16$	$0 \cdot 86$
$5 \cdot 0 \times 10^4$	$0 \cdot 67$	~ 3
$1 \cdot 0 \times 10^5$	$0 \cdot 73$	~ 4
$2 \cdot 5 \times 10^5$	$0 \cdot 76$	~ 4
Observed	$0 \cdot 166$	$0 \cdot 79$

Half-widths of the emergent Hα lines for various physical conditions may be calculated from a knowledge of the quantities occurring in (18) and are shown in Table 11, together with central intensities, for $N_0 = 5 \times 10^{11}$ cm.$^{-3}$ and $\beta = 6 \times 10^{-9}$ cm.$^{-1}$.

When the optical depth in Hα is very small, the emergent intensity will be given approximately by equation (23) which, from equation (26), represents a line of half-width about 4 Å, and central intensity about $0 \cdot 75\ J_w$.

On the other hand, when the central intensity is relatively low the photo-
spheric component contributes to the emergent radiation as soon as the chromo-
spheric optical depth becomes small enough to allow it. In normal circumstances
this occurs about 1 Å, or even less, from the centre and consequently the half-
width in this case is controlled by both photosphere and chromosphere, being
of the order of $0 \cdot 5$ Å. This accounts for the very great differences among
the values of the half-widths shown in Table 11.

The mean central intensity and half-width from the above four sets of solar
observations are given at the bottom of the corresponding column of Table 11.
The agreement between observed values and those calculated for $T = 2 \cdot 5 \times 10^4$ °K.
is probably only fortuitous, owing to the approximate methods involved in the
calculations and the uncertainty in the data. It would seem, however, that
observations are compatible with temperatures up to about $3 \cdot 5 \times 10^4$ °K. in
those regions of the chromosphere which give rise to the observed Hα radiation.

IX. Non-coherent Scattering

The results obtained above for the contours and central intensities of the
Lyman and Hα lines are based explicitly on the assumption of coherent scattering
whose intensity is independent of direction. We shall now obtain the form
of the contour in Lα and Hα with a simple model of non-coherent scattering
whose intensity is still uniform with angle. We assume that non-coherency is
introduced by a Doppler shift in frequency associated with a Maxwellian velocity
distribution at the kinetic temperature T, implying that the finite width of the
line is due to thermal velocities alone. Doppler shifts are not uniformly
distributed with angle, but for simplicity the angular variation is neglected.
This should give reasonable results at frequencies not too far from the centre of
the line.

With coherent scattering a quantum of frequency ν' which is absorbed and
subsequently scattered is re-emitted with the same frequency. In the present
case, however, the probability that the frequency of the scattered quantum will
lie between ν and $\nu + d\nu$ is given by

$$p(\nu)d\nu = \sqrt{\gamma/\pi} \exp[-\gamma(\nu - \nu_0)^2]d\nu, \quad \dots \dots \dots \dots \quad (27)$$

where

$$\gamma = \sqrt{\frac{M}{2kT}} \frac{c}{\nu_0},$$

M being the mass of the hydrogen atom. Then the total energy emitted per
second from unit volume into unit solid angle per unit frequency range is

$$E_\nu = \frac{(1-\lambda)\sqrt{\gamma/\pi} \exp[-\gamma(\nu - \nu_0)^2] \int_0^\infty J_{\nu'}\alpha_{\nu'}d\nu'}{4\pi} + \varepsilon_\nu,$$

$$\dots \dots \dots \dots \dots \dots \quad (28)$$

where ε_ν is the " true " emission of the medium.

For simplicity we take $J_{\nu'}$ to be constant for all frequencies and equal to J_0,
the value at the centre of the line, an assumption which makes little difference
to the value of the integral.

Since $\alpha_\nu = \alpha_0 \exp[-\gamma(\nu - \nu_0)^2]$ over the region of Doppler absorption, it follows that equation (28) then becomes

$$E_\nu = \frac{(1-\lambda)\,\exp[-\gamma(\nu-\nu_0)^2]J_0\alpha_0}{4\pi} + \varepsilon_\nu, \quad \ldots\ldots (29)$$

and

$$B_\nu = \frac{E_\nu}{\alpha_\nu} = \frac{(1-\lambda)J_0}{4\pi} + \frac{\varepsilon_\nu}{\alpha_\nu}. \quad \ldots\ldots\ldots\ldots\ldots (30)$$

Making use of Eddington's approximation we may write the equation of radiative transfer in the form (see, e.g. Rosseland 1936)

$$\frac{1}{3}\frac{\partial^2 J_\nu}{\partial \tau_\nu^2} = J_\nu - 4\pi B_\nu, \quad \ldots\ldots\ldots\ldots\ldots (31)$$

which by equation (36) gives

$$\frac{1}{3}\frac{\partial^2 J_\nu}{\partial \tau_\nu^2} = J_\nu - (1-\lambda)J_0 - \frac{4\pi\varepsilon_\nu}{\alpha_\nu}. \quad \ldots\ldots\ldots\ldots (32)$$

Assuming, as before, that λ is uniform throughout the atmosphere, and that $4\pi\varepsilon/\alpha\lambda$ is constant or varies only linearly with optical depth, the solution of equation (32) for the central intensity is

$$J_0 = \frac{4\pi\varepsilon}{\alpha\lambda} + a \exp(\sqrt{3\lambda}\tau) + b \exp(-\sqrt{3\lambda}\tau), \quad \ldots\ldots (33)$$

where subscripts in ν_0 have been omitted on the right-hand side.

Substituting J_0 from (33) we find

$$\frac{1}{3}\frac{\partial^2 J_\nu}{\partial \tau_\nu^2} = J_\nu - \frac{4\pi\varepsilon}{\alpha\lambda} - (1-\lambda)[a\exp(\sqrt{3\lambda}k\tau_\nu) + b\exp(-\sqrt{3\lambda}k\tau_\nu)],$$
$$\ldots\ldots\ldots\ldots\ldots (34)$$

where $\tau_0 = k\tau_\nu$, that is, $k = \exp\gamma(\nu - \nu_0)^2$.

The solution of (34) follows simply, giving

$$J_\nu = \frac{4\pi\varepsilon}{\alpha\lambda} + A\exp(\sqrt{3}\tau_\nu) + B\exp(-\sqrt{3}\tau_\nu)$$
$$+ \frac{1-\lambda}{1-\lambda k^2}[a\exp(\sqrt{3\lambda}k\tau_\nu + b\exp(-\sqrt{3\lambda}k\tau_\nu)]. \quad \ldots\ldots\ldots (35)$$

The relevant boundary conditions are the same as in the case of coherent scattering and are applicable to each frequency. For Lyman line radiation, evaluation of the constants a, b, A, and B gives the following expression for the emergent intensity

$$\frac{J_\nu}{4\pi\varepsilon/\alpha\lambda} = 1 - R_1 - R_2, \quad \ldots\ldots\ldots\ldots (36)$$

where

$$R_1 = [1 + \exp(-2\tau_1{}^\nu\sqrt{3})]$$
$$\times \left[1 - \frac{\dfrac{1-\lambda}{1-\lambda k^2}\left\{\dfrac{1+2k\sqrt{\lambda/3} + (1-2k\sqrt{\lambda/3})\exp(-2\tau_1{}^\circ\sqrt{3\lambda})}{1+2\sqrt{\lambda/3} + (1-2\sqrt{\lambda/3})\exp(-2\tau_1{}^\circ\sqrt{3\lambda})}\right\}}{1+2/\sqrt{3} + (1-2/\sqrt{3})\exp(-2\tau_1{}^\nu\sqrt{3})}\right]$$

and

$$R_2 = \frac{1-\lambda}{1-\lambda k^2}\left[\frac{1+\exp(-2\tau_1{}^\circ\sqrt{3\lambda})}{1+2\sqrt{\lambda/3} + (1-2\sqrt{\lambda/3})\exp(-2\tau_1{}^\circ\sqrt{3\lambda})}\right].$$

For Hα, the expression for J_ν is very unwieldy but follows in a straight-forward manner on application of the boundary conditions.

The Lα and Hα intensities at various distances from the line centres are shown in Tables 12 and 13. The calculations are made for an atmosphere with $T = 2 \cdot 5 \times 10^4$ °K. and $\beta = 6 \times 10^{-9}$ cm.$^{-1}$. $N_0 = 5 \times 10^{11}$ cm.$^{-3}$ for Lα and

TABLE 12

EMERGENT INTENSITY ACROSS Lα

$T = 2 \cdot 5 \times 10^4$ °K., $\beta = 6 \times 10^{-9}$ cm.$^{-1}$, $N_0 = 5 \times 10^{11}$ cm.$^{-3}$

Δλ (Å)	Non-coherent Scattering	Coherent Scattering
0	0·018	0·018
0·05	0·022	0·018
0·10	0·041	0·018
0·15	0·156	0·011
0·17	0·298	0·004
0·20	0·442	0·001
0·225	0·419	0
0·25	0·149	0
0·275	0·025	0
0·30	0·002	0

2×10^{11} cm.$^{-3}$ for Hα. The double entries in the second and third columns of Table 13 represent the chromospheric (left) and photospheric (right) components of the emergent radiation. With Hα, non-coherent scattering over a Doppler

TABLE 13

EMERGENT INTENSITY NEAR THE CENTRE OF Hα

$T = 2 \cdot 5 \times 10^4$ °K., $\beta = 6 \times 10^{-9}$ cm.$^{-1}$, $N_0 = 2 \times 10^{11}$ cm.$^{-3}$

Δλ (Å)	Non-coherent Scattering		Coherent Scattering	
0	0·158	0·008	0·158	0·008
0·14	0·161	0·009	0·158	0·012
0·29	0·166	0·019	0·152	0·038
0·43	0·164	0·072	0·133	0·127
0·57	0·136	0·266	0·092	0·320

profile is assumed in both the chromosphere and photosphere. This restriction—to a Doppler profile—means that we are only able to compare the profiles near the centre of the line as the two cases are clearly not comparable in the wings. It will be seen that the contour of the Lyman line is considerably affected by the introduction of non-coherent scattering; the Hα contours are practically

identical for coherent and non-coherent scattering in the particular case considered.

With the Lyman lines, the introduction of non-coherent scattering results in a broadening, and formation of an M-shaped contour in the emergent line, as may be seen from Table 12. The contours at the base of the atmosphere are similar for both types of scattering, each being flat-topped. The non-coherent profile is, however, still the broader of the two. It is clear that the computed central intensities of all lines will be the same as those obtained on the assumption of coherent scattering since, by replacing J_ν with J_0 in the integral of equation (34), we have recovered the equation of radiative transfer for the case of coherent scattering.

The very great difference in behaviour between the Lα and Hα contours for coherent and non-coherent scattering is due primarily to the differences in τ for the two lines. In the case considered, the absorption coefficient for Lα radiation is of the order of 100 times that for Hα, at the centre of the lines. Lα quanta are virtually trapped at the centre of the line, but may escape freely in the wings when their frequencies are redistributed by non-coherent scattering.

The variation of the shape of the contour with kinetic temperature may readily be seen. At temperatures greater than 5×10^4 °K., the optical depths in Hα and Lα for normal values of β and N_0 are so small that chromospheric scattering is unimportant. At temperatures below 5×10^4 °K., non-coherent scattering broadens the contour, the M-shaped profile appearing when the optical depth becomes fairly large.

X. Discussion

It is rather difficult to see to what extent the results obtained in this paper may be applied to the Sun. The only direct observational evidence which is available for check is the Hα contour, in particular the central intensity and half-width. The computed values of these quantities are based on an approximate solution of the equation of radiative transfer, whose terms are calculated with reasonable values of N_e, variations of which do not in general greatly affect the computed central intensity.

In the absence of a detailed knowledge of the physical structure of the chromosphere it seems difficult to improve on the approximations, or to estimate the error involved. It seems, however, that on the models adopted here the results quoted for central intensities should be correct to rather better than a factor of 2 at the lower temperatures, and so we should be justified on the basis of the computed central intensities in saying that the kinetic temperature of the regions responsible for the observed Hα radiation lies somewhere in the range below $3 \cdot 5 \times 10^4$ °K.

It is possible to make some inferences on the chromospheric temperature from observations of the Hα profile. The observed rapid initial increase in intensity as we move away from the centre of the line must almost certainly be attributed to the effect of photospheric radiation penetrating the chromosphere, rather than one due solely to variations in the composition of

the chromosphere, since calculations indicate that the quantity $4\pi\varepsilon/\alpha\lambda$ does not vary sufficiently with electron concentration for this latter effect to be significant. The most important parameter in determining the contour at a given electron temperature is the total optical depth, which in turn depends on N_0, the electron concentration at the base. If this is large enough, the intensity of the emergent line will be effectively constant for some distance from the centre, as may be seen from Figure 5. The value of N_0 for $T = 2\cdot5 \times 10^4$ °K., which best fits the observations, is 2×10^{11} cm.$^{-3}$, and this is the value adopted in Figures 4 and 5. For $T = 10^4$ °K. and $N_0 = 2 \times 10^{11}$ cm.$^{-3}$, $\sqrt{3}\lambda\tau_1° \cong 30$. This is rather too large to allow the photosphere to contribute at all before $\lambda = 0\cdot5$ Å. For this temperature the central intensity is rather low, viz. $0\cdot09$, suggesting that, if it possesses a temperature in the range discussed here, i.e. 10^4 °K. or more, the $H\alpha$-emitting region of the chromosphere has a temperature of at least $1\cdot5 \times 10^4$ °K.

XI. ACKNOWLEDGMENTS

XII. REFERENCES

BATES, D. R., FUNDAMINSKY, A., LEECH, J. W., and MASSEY, H. S. W. (1950).—*Phil. Trans.* **243** : 93.

BOHR, N., PEIERLS, R., and PLACZEK, G. (1949).—Quoted by Mott. N. F., and Massey, H. S. W. (1949).—" The Theory of Atomic Collisions." 2nd Ed. p. 135. (Oxford Univ. Press.)

TEN BRUGGENCATE, P., GOLLNOW, H., GUNTHER, S., and STROHMEIER, W. (1949).—*Z. Astrophys.* **26** : 51.

CHANDRASEKHAR, S., and BREEN, F. H. (1946).—*Astrophys. J.* **104** : 430.

EVANS, D. S. (1940).—*Mon. Not. R. Astr. Soc.* **100** : 156.

GIOVANELLI, R. G. (1949).—*Mon. Not. R. Astr. Soc.* **109** : 298.

MINNAERT, M. (1927).—*Z. Phys.* **45** : 610.

MIYAMOTO, S. (1951a).—*Publ. Astr. Soc. Japan* **2** : 102.

MIYAMOTO, S. (1951b).—*Publ. Astr. Soc. Japan* **2** : 113.

PURCELL, E. M. (1952).—*Astrophys. J.* (in press).

ROSSELAND, S. (1936).—" Theoretical Astrophysics." p. 112. (Oxford Univ. Press.)

THACKERAY, A. D. (1935).—*Mon. Not. R. Astr. Soc.* **95** : 293.

THOMAS, R. N. (1948).—*Astrophys. J.* **108** : 142.

THOMAS, R. N. (1949).—*Astrophys. J.* **109** : 480.

UNSOLD, A. (1938).—" Physik der Sternatmospharen." p. 189. (J. Springer : Berlin.)

WHITE, H. E. (1934).—" Introduction to Atomic Spectra." p. 435. (McGraw-Hill : New York.)

NEUTRONS EMITTED IN THE DISINTEGRATION OF BERYLLIUM BY DEUTERONS

By A. J. Dyer* and J. R. Bird*

Summary

Neutrons obtained by bombarding a thick beryllium target with 600 keV. deuterons were studied by the photographic plate technique at angles of emission of 0, 90, and 150°. Five previously reported neutron groups were confirmed giving energy levels of B^{10} at 0·73, 1·75, 2·20, and 3·64 MeV. An additional group corresponding to an energy level of 2·85 MeV. was observed and the evidence for regarding these as genuine beryllium neutrons is discussed. The Q-value for the ground state transition is found to be $4·35 \pm 0·02$ MeV. which agrees with the value $4·36 \pm 0·04$ deduced from the most recent masses.

I. INTRODUCTION

The energy spectrum of the neutrons obtained from deuteron bombardment of beryllium was first investigated with cloud chambers by Bonner and Brubaker (1936) and Staub and Stephens (1939), and their results indicated the presence of energy levels in B^{10} at 0·7, 2·2, and 3·6 MeV. An early demonstration of the photographic plate technique for neutron detection gave similar results (Powell 1943). Whitehead and Mandeville (1950) using Ilford C_2 emulsions showed the presence of a further level at 5·1 MeV.

Using a bombarding energy of 3·4 MeV. and a thin beryllium target, Ajzenburg (1951) demonstrated that neutrons hitherto attributed to the D–D reaction were in fact a genuine beryllium group, and reported levels at 4·8 MeV. and higher energies. Pruitt, Hanna, and Swartz (1952) obtained similar results.

A study of the reaction $Li^7(\alpha,n)B^{10}$ by Haxel and Stuhlinger (1939) indicated levels at 0·8, 1·3, and 2·1 MeV.

The γ-radiation from the excited B^{10} nucleus, has been studied by Rasmussen, Hornyak, and Lauritsen (1949) who suggested levels at 0·412, 0·713, 1·017, 1·42, 2·14, 2·86, 3·58, and 5·16 MeV. and commented on the possible existence of an equally spaced set of levels 0·713 MeV. apart. Not all of these levels, however, are necessary to explain the observed γ-radiation.

The purpose of the present experiment was to investigate further the possibilities of weak neutron groups from this reaction and verify the existence of the 1·79 MeV. level reported by Ajzenburg.

* Physics Department, University of Melbourne.

II. Experimental Technique

A thick target of beryllium metal 2 mm. in diameter mounted on a thin copper disc was bombarded for 100 μAh. by a 600 keV. deuteron beam. Preliminary investigations using a thin evaporated target gave insufficient yield for the detection of weak neutron groups.

Quarter sections of 50μ Ilford C_2 plates were placed at various angles to the incident deuteron beam and with the surface of the emulsion in the direction of the neutron flux. The plates were supported by fine wires inside a thin-walled brass doughnut at a distance of 12 cm. from the target. The two straight edges of each quarter section were used to locate the plates to within 0·5° of the required angles.

Prior to use the plates were vacuum dried overnight and a phosphorus pentoxide drying agent was used to prevent absorption of moisture by the emulsion during the exposure period.

III. Measurements

Measurements were made on plates exposed at 0, 90, and 150° to the deuteron beam over areas subtending ±1° at the target. The range distribution of the proton recoil tracks was measured at two magnifications covering the energy intervals 0–3 and 2–5 MeV. respectively. Proton recoils were selected which made an angle of not more than 5° with the neutron direction. This was effected by accepting tracks whose initial direction lay within 5° lines marked on an eyepiece graticule, and whose angle of dip in the unprocessed emulsion was less than 5°. Taking into account the known depth of focus of the objectives used and the emulsion shrinkage factor (Martin 1949) the restriction on the angle of dip requires that the beginning of the tracks be in focus for 10 and 20μ at high and low magnification respectively.

The eyepiece scales used for range measurements were calibrated against a Grayson ruling, which had been checked with a Hilger travelling microscope, and all range determinations are considered to be accurate to within 0·3 per cent.

A total of 4200 tracks which had their whole length within the emulsion was recorded.

The observations were plotted as energy distributions in intervals of 100 keV. using a range-energy curve deduced from the work of El-Bedewi (1951), Rotblat (1951), and Dyer (1952).

These distributions were then corrected for the variation with energy of the neutron-proton scattering cross section (Bailey *et al.* 1946 ; Lampi, Freier, and Williams 1950) and the loss of tracks from the emulsion (Richards 1941).

At each angle the observations for the two magnifications were then normalized using the overlapping regions. The angular distribution of the low energy peak was measured separately under constant conditions of observation and used to adjust the ordinate scale of the three curves to express the actual variation of neutron intensity (Fig. 1).

IV. Analysis of Results

In order to obtain the " mean thin target " energy for each neutron group it was necessary to make a thick target correction and the results for each peak
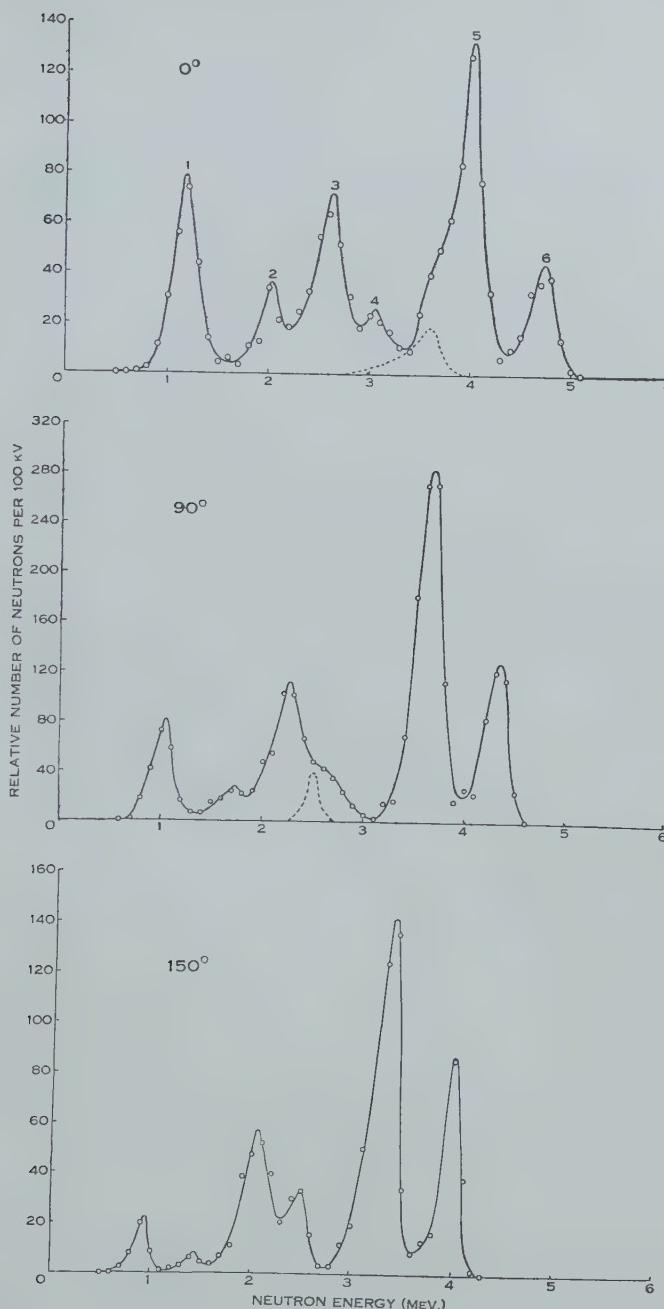
Fig. 1.—Neutron spectra.

were put into the integral form. This is a straightforward procedure for well-separated peaks (e.g. peak 6) but where overlapping peaks occur a suitable shape for each must be deduced.

D

Livesey and Wilkinson (1948) showed that a differential thick target distribution can be expressed in the analytic form

$$f(y) = e^{-2\lambda y} \left\{ 1 - \mathrm{erf}\left(\lambda\sigma - \frac{y}{\sigma}\right) \right\},$$

where the abscissa y is measured from the mean thin target energy. This expression is obtained by superposing a Gaussian straggling distribution of
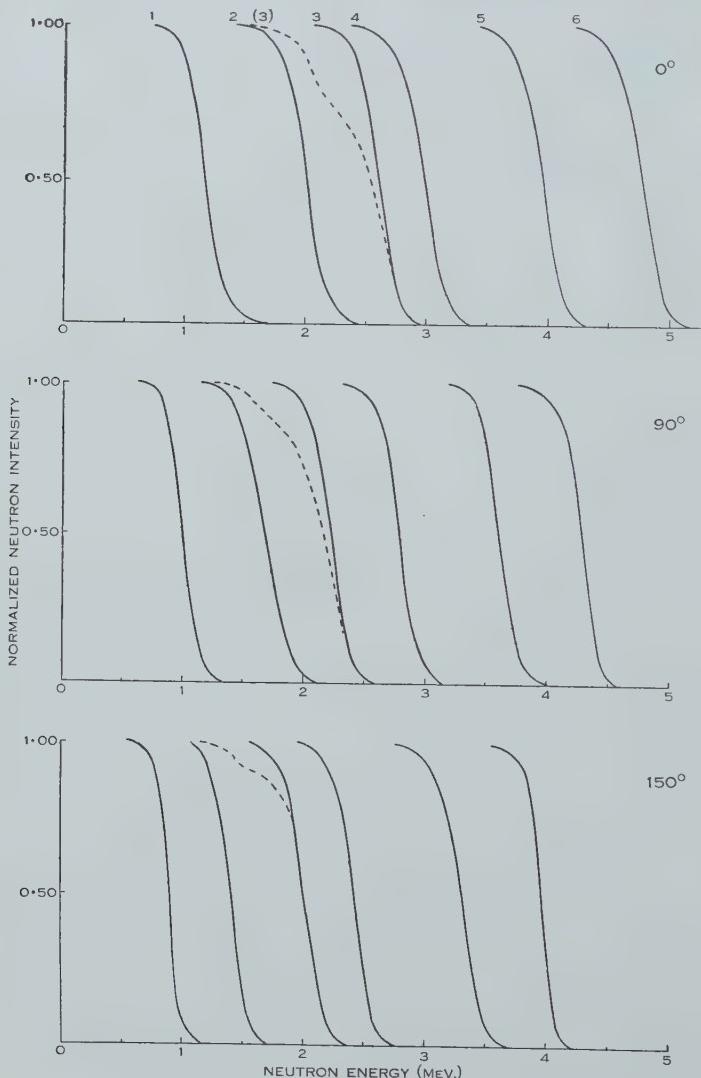
Fig. 2.—Normalized integral curves.

variance σ^2 on an exponential function $e^{-2\lambda y}$, which describes the theoretical neutron spectrum deduced from the excitation function. A value of σ of about $0 \cdot 2$ MeV. was found to fit the observed straggling and two sets of theoretical

integral curves were drawn with σ equal to $0 \cdot 16$ and $0 \cdot 24$ MeV. for a range of λ of 1–9 MeV.$^{-1}$.

From the dynamics of the reaction it is expected that λ will be independent of energy for neutrons observed at $90°$ and will slowly decrease with energy at $0°$ and increase at $150°$. Making use of this information together with the theoretical curves, the overlapping peaks were resolved and values of λ and σ found which fitted the observations in every case. In this way a set of curves for each spectrum was obtained and the total number of tracks in each curve was normalized to unity (Fig. 2).

The mean thin target energy can be obtained by the method of Livesey and Wilkinson in which the extrapolated energy from the theoretical curve is applied as a correction to the experimental extrapolated energy. This requires the drawing of the steepest gradient which is an inherently difficult procedure. An alternative method is to determine the fraction of the total number of particles having an energy greater than the mean thin target energy from the theoretical curve, and to find the energy for which this fraction occurs on the experimental curve. In most cases the results obtained by the two methods were nearly identical and a mean value was accepted. For those cases in which a discrepancy occurred it was felt that the second method gave a more reliable result. The Q-values obtained from these energy values were tabulated and a mean determined by weighting each according to the square root of the number, N, of tracks represented (Table 1).

TABLE 1

Q-VALUES FOR INDIVIDUAL NEUTRON GROUPS AT DIFFERENT ANGLES OF EMISSION

The errors expressed are probable errors

Peak No. \ Angle	1	$\sqrt{N}$	2	$\sqrt{N}$	3	$\sqrt{N}$	4	$\sqrt{N}$	5	$\sqrt{N}$	6	$\sqrt{N}$
$0°$	$0 \cdot 70$	18	$1 \cdot 52$	10	$2 \cdot 15$	19	$2 \cdot 58$	12	$3 \cdot 58$	16	$4 \cdot 35$	9
$90°$	$0 \cdot 72$	13	$1 \cdot 51$	8	$2 \cdot 10$	14	$(2 \cdot 70)$	5	$3 \cdot 61$	19	$4 \cdot 35$	$11 \cdot 5$
$150°$	$0 \cdot 75$	7	$1 \cdot 39$	4	$2 \cdot 17$	14	$2 \cdot 63$	11	$3 \cdot 65$	25	$4 \cdot 35$	11
Mean	$0 \cdot 71 \pm 0 \cdot 02$		$1 \cdot 50 \pm 0 \cdot 03$		$2 \cdot 15 \pm 0 \cdot 02$		$2 \cdot 60 \pm 0 \cdot 02$		$3 \cdot 62 \pm 0 \cdot 02$		$4 \cdot 35 \pm 0 \cdot 02$	

The values tabulated correspond to the peaks numbered 1 to 6 in Figure 1. The dotted peaks are attributed to D–D neutrons and give a mean Q-value of $3 \cdot 25 \pm 0 \cdot 06$ MeV. which is in excellent agreement with the result of Livesey and Wilkinson (1948). The Q-value obtained for neutrons observed at $90°$ for peak 4 was not used in calculating the mean value for this group since it depended very strongly on the intensity allotted to the D–D group immediately next to it. The Q-value of the ground-state transition $4 \cdot 35 \pm 0 \cdot 02$ MeV. agrees with the value of $4 \cdot 36 \pm 0 \cdot 04$ MeV. deduced from the most recent masses (Li *et al.* 1951).

V. Discussion

Peaks 1, 3, 5, and 6 are well-established beryllium neutron groups and peak 4 confirms the existence of the doublet reported by Ajzenburg (1951). Peak 2 has not been previously observed but since measurements were taken at three angles of emission, it is possible to check that it is a genuine beryllium group.

It can be seen from Table 1 that the Q-values for peak 2 obtained at 0 and 90° are in excellent agreement, while the value at 150° is still statistically significant. This can be investigated further by plotting the energy shifts E_0-E_{90} and $E_{90}-E_{150}$ against E_0 and E_{150} respectively, for each neutron group (Fig. 3).

It can be shown from the dynamics of a nuclear reaction that

$$E_0-E_{90} = 2 \cdot 19 \sqrt{E_0} \;/(M+2)$$
$$E_{90}-E_{150} = 1 \cdot 90 \sqrt{E_{150}}/(M+2),$$

where M is the mass of the target nucleus. These expressions have been included in Figure 3 for a range of values of M. The experimental points lie about the
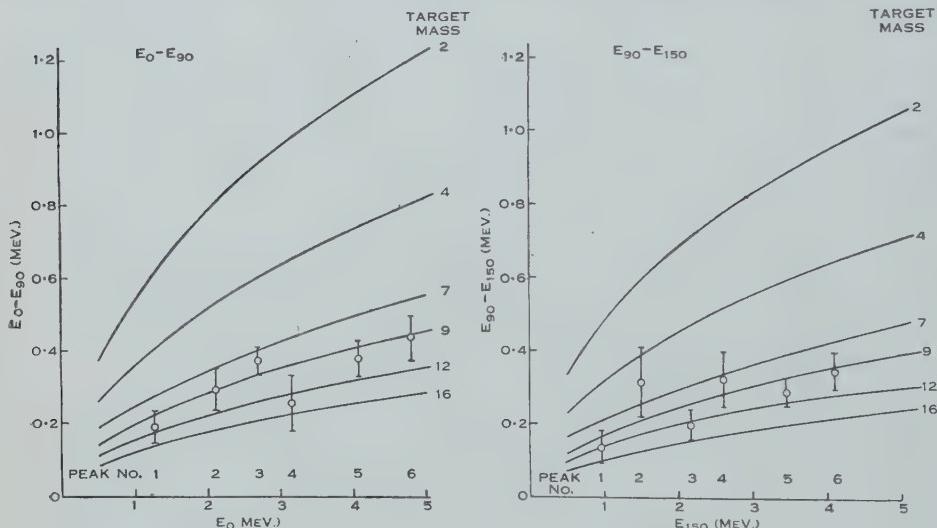
Fig. 3.—Energy shift curves.

curve $M=9$. Peak 2 could only arise from a reaction involving a target nucleus of mass <14, and no such neutron reactions giving these energies at the three angles are known.

If it is assumed that peak 2 does not exist, the integral curves for peak 3 are given by the dotted lines in Figure 2. The deviations of these curves from the required shape is far greater than can be explained by statistical fluctuations.

At a bombarding energy of $3 \cdot 41$ MeV. and using a thin target Ajzenburg (1951) reports that no neutrons were observed corresponding to this group. This author states that at a high bombarding energy one would expect that a large number of levels of the compound nucleus covering a wide range of angular momenta and of either parity would be involved in the reaction. Neutron groups should thus be observed corresponding to all existing levels of B^{10}. It must be

pointed out, however, that the relative intensities of the various groups observed indicate that a considerable degree of competition between the various modes of neutron emission is still taking place. The possibility that a weak neutron group would escape detection under these circumstances, but would be observed for lower bombarding energies with a thick target is, therefore, not entirely ruled out.

It is interesting to note that, in the work of Staub and Stephens (1939) and Powell (1943) using a low bombarding energy and a thick target, there is in fact some evidence of a weak neutron group in this region which has been previously ignored. Since both these observations were made at an angle of 90° to the incident deuterons, the expected intensity would be small.

VI. Energy Levels in B^{10}

The Q-values in Table 1 lead to the energy levels of B^{10} presented in Figure 4, together with the corresponding values found by Ajzenburg (1951) and Pruitt,
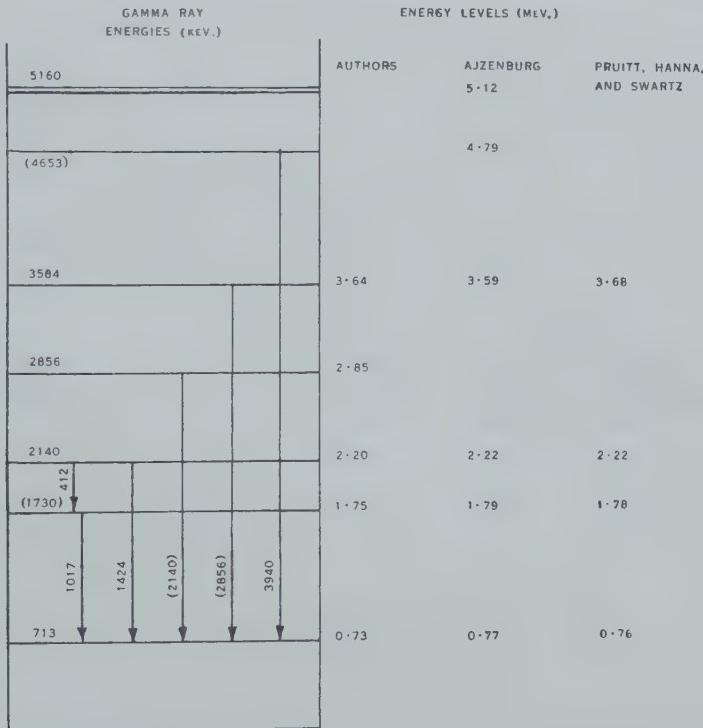

Fig. 4.—Energy levels in B^{10}, showing possible γ-ray transitions in addition to ground-state transitions.

Hanna, and Swartz (1952). It is reasonable to assume that the energies of the lower levels will be given more accurately by γ-ray determinations. The values obtained by Rasmussen, Hornyak, and Lauritsen (1949) are also shown in Figure 4.

The level scheme is essentially that proposed by Ajzenburg. All neutron measurements give consistently higher energy values than the γ-ray results but the agreement appears to be better in our case. The additional level at 2·86 MeV. gives a consistent explanation of the observed γ-rays as an alternative to the scheme of Ajzenburg. This level, together with the one at 1·3 MeV. reported by Haxel and Stuhlinger (1939) from the reaction $Li^7(\alpha,n)B^{10}$, again raises the possibility of equally spaced levels suggested by Rasmussen, Hornyak, and Lauritsen. No neutrons were observed, however, corresponding to a level at 1·4 MeV., and some explanation would be required for the raised levels at 1·73 and 4·65 MeV.

It is not possible to check the validity of the proposed level scheme by a comparison of neutron and γ-ray relative intensities, and no confirmatory evidence is available from other reactions giving rise to low-lying levels in B^{10}. Further work on neutron resonances and γ-rays is therefore desirable to check the existence and position of the energy levels.

VII. Acknowledgments

Our thanks are due to Professor L. H. Martin for his active interest in the experiment. We also desire to acknowledge the assistance of Mr. J. G. Campbell in making the exposures.

VIII. References

Ajzenburg, F. (1951).—*Phys. Rev.* **82** : 43.
Bailey, C. L., Bennett, W. E., Bergstralh, T., Nuckolls, R. G., Richards, H. T., and Williams, J. H. (1946).—*Phys. Rev.* **70** : 583.
Bonner, T. W., and Brubaker, W. M. (1936).—*Phys. Rev.* **50** : 308.
Dyer, A. J. (1952).—*Aust. J. Sci. Res.* A **5** : 104.
El-Bedewi, F. A. (1951).—*Proc. Phys. Soc. Lond.* A **64** : 584.
Haxel, O., and Stuhlinger, E. (1939).—*Z. Phys.* **114** : 178.
Lampi, E. E., Freier, G. D., and Williams, J. H. (1950).—*Phys. Rev.* **80** : 853.
Li, C. W., Whaling, W., Fowler, W. A., and Lauritsen, C. C. (1951).—*Phys. Rev.* **83** : 512.
Livesey, D. L., and Wilkinson, D. H. (1948).—*Proc. Roy. Soc.* A **195** : 123.
Martin, S. L. (1949).—*Aust. J. Sci. Res.* A **2** : 389.
Powell, C. F. (1943).—*Proc. Roy. Soc.* A **181** : 344.
Pruitt, J. S., Hanna, S. S., and Swartz, C. D. (1952).—*Phys. Rev.* **87** : 534.
Rasmussen, V. K., Hornyak, W. F., and Lauritsen, T. (1949).—*Phys. Rev.* **76** : 581.
Richards, H. T. (1941).—*Phys. Rev.* **59** : 796.
Rotblat, J. (1951).—*Nature* **167** : 550.
Staub, H., and Stephens, W. E. (1939).—*Phys. Rev.* **55** : 131.
Whitehead, W. O., and Mandeville, C. E. (1950).—*Phys. Rev.* **77** : 732.

ANGULAR CORRELATION BETWEEN α–PARTICLES AND γ–RAYS IN THE Be⁹(d,α)Li⁷*γLi⁷ REACTION

By R. G. Uebergang* and N. W. Tanner*

[*Manuscript received November 28, 1952*]

Summary

The angular correlation of the α-particles and γ-rays in the Be⁹(d,α)Li⁷*γLi⁷ reaction has been measured. The results show no significant departure from isotropy. The most plausible explanation is that the lowest excited state of Li⁷ has spin $\frac{1}{2}$.

I. Introduction

Attempts to assign a spin value which is above question to the lowest excited state of Li⁷ have met with unexpected difficulties.

Early experiments based on the B¹⁰(n,α)Li⁷*γLi⁷,Li⁷ branching ratio indicated a spin of 5/2 (Inglis 1948 ; Feld 1949) which was difficult to interpret theoretically.

Angular correlation measurements between the particle associated with the transition leading to the Li⁷ excited state and the γ-radiation emitted as this state decays have been carried out for three nuclear reactions. In general an anisotropic correlation should occur for Li⁷* spin $> \frac{1}{2}$ but it should be isotropic for spin $= \frac{1}{2}$. For example, the angular correlation between the α-particle and γ-ray from the B¹⁰(n,α)Li⁷*γLi⁷ reaction has been found by Rose and Wilson (1950) to be spherically symmetric, which is consistent with a Li⁷* spin value $\frac{1}{2}$. However, the isotropy observed in this experiment can also be explained by a spin value of 5/2 and a possible but unlikely admixture of electric quadrupole and magnetic dipole γ-radiation (Inglis 1951 ; Class and Hanna 1952).

The spin assignment of $I = \frac{1}{2}$ for Li⁷* is also supported by the measurement of the angular distribution of γ-rays from the inelastic scattering of protons by Li⁷ (Littauer 1950), but it has been suggested by Inglis (1951) that the proximity of the bombarding proton energy to a resonance in the compound nucleus, which may have zero total angular momentum, renders this experiment inconclusive.

Newton (1951) has pointed out that the chance cancellation of asymmetry terms which is possible in the B¹⁰(n,α)Li⁷*γLi⁷ reaction is unlikely to occur for all other reactions leading to the excited state. Four groups measured the p-γ angular correlation in Li⁶(d,p)Li⁷*γLi⁷ almost simultaneously (Class and Hanna 1951, 1952 ; Newton 1951 ; Burke and Risser 1952 ; Phillips, Heydenburg, and Cowie 1952) ; all reported spherical symmetry within the accuracy of their measurements. Class and Hanna have shown that in this experiment the dipole, quadrupole radiation mixture cannot cause cancellation of asymmetry terms, possible in the B¹⁰(n,α)Li⁷*γLi⁷ reaction. Unfortunately

* Physics Department, University of Melbourne.

the populations of the magnetic substates of the Li^{7*} nuclei cannot be calculated because of the multiple values of orbital angular momenta of the deuterons and protons involved and because of channel spin degeneracy of the kind discussed by Biedenharn, Arfken, and Rose (1951) for the $B^{11}(p,\gamma,\gamma)C^{12}$ reaction. There is, therefore, the possibility that these populations may be nearly equal and so lead to isotropy.

Because of these doubts it was thought worth obtaining information from a further reaction and therefore a measurement of the α-γ angular correlation associated with $Be^9(d,\alpha)Li^{7*}\gamma Li^7$ has been carried out. Calculations based on the method of Biedenharn, Arfken, and Rose (1951) indicate that, for a Li^7 excited state spin value of $5/2$ and plausible angular momenta, the correlation may be expected to be of the order of $0\cdot20$ except for a chance cancellation similar to that discussed above.

II. EXPERIMENTAL

The 750 kV. electrostatic generator of the Physics Department, University of Melbourne, was used to supply 400 keV. deuterons.

The deuterons were analysed in a 90° sector magnet and thence passed horizontally through a 3 ft. length of 1 in. diameter tubing to reach the target chamber. The cylindrical target chamber (Fig. 1) was mounted with its axis vertical and coincident with the centre of rotation of the γ-counter, and rigidly attached to the central bearing carrying the γ-counter so that no relative to and fro motion was possible. Approximately $2\cdot5$ cm. from the target the beam was defined by a $0\cdot32$ cm. diameter aperture. A disk of beryllium $0\cdot004$ in. thick was used as target, mounted in a vertical plane on the axis of the chamber in such a manner that the angle between the deuteron beam and the target plane could be varied to check the absorption of γ-rays in the target.

α-Particles from the reaction were detected by a scintillation counter mounted at 152° to the deuteron beam direction in the horizontal plane. Silver-activated zinc sulphide was selected as the phosphor for this counter as it is very insensitive to γ-rays and neutrons. The powder was deposited uniformly on a glass disk by a method previously described (Hirst and Uebergang 1951) and this glass disk was cemented to a glass light pipe which led the scintillations through the target chamber wall to a Mazda 27 M1 photomultiplier tube. To exclude light and avoid counting scattered deuterons the screen was covered by four layers of aluminium leaf, each of 1 mm. air equivalent stopping power. This thickness of aluminium leaf is sufficient to prevent counts from recoil Li^{7+++} nuclei from the reaction.

γ-Rays were detected, after passing through the uniform $\frac{1}{32}$ in. thick wall of the target chamber, by a scintillation counter consisting of a NaI(Tl) crystal and a 5359 EMI photomultiplier. The 1 in. diameter by 1 in. long crystal was immersed in oil in a close-fitting aluminium container which was seated on a rubber " O " ring on the flat photomultiplier end. An oil seal was provided by fitting tight rubber sleeving over the container and photomultiplier.

To improve the energy resolution both the sodium iodide crystal surface and the inner surface of the container were polished. The counter could be

moved concentrically about the target to make any angle from 55 to 275° to the α-particle detector.

The solid angle subtended by each counter to the target was approximately 0·14 sterad., that is, approximately 12° half angle.

In addition to the reaction under study the bombardment of beryllium by deuterons gives rise to several competing reactions emitting protons, tritons, and neutrons and associated γ-rays. To these various emissions a small contribution from the bombardment of oxygen and carbon contamination may be
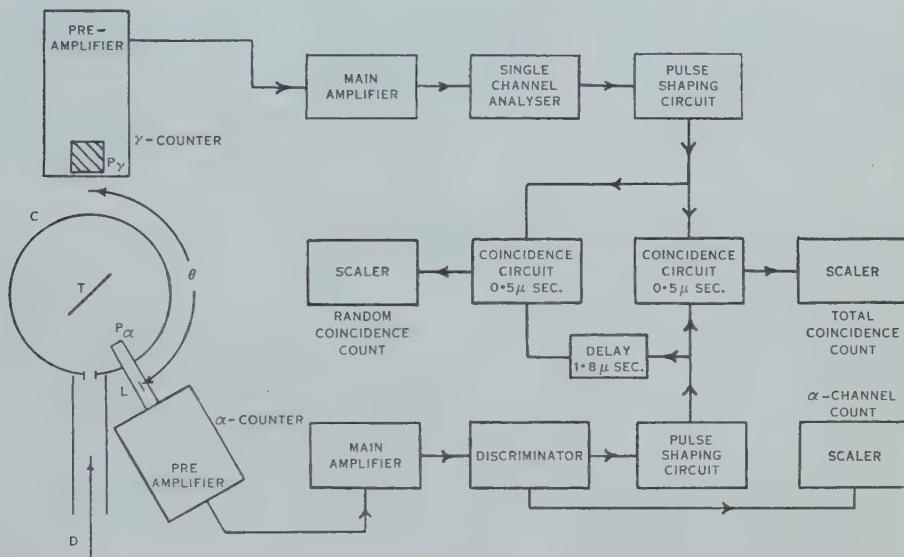
Fig. 1.—Block schematic of counting arrangement. D, deuteron beam; C, target chamber; P_γ, γ-counter phosphor NaI(Tl); P_α, α-counter phosphor ZnS (Ag); L, light pipe; θ, correlation angle.

added (Hornyak *et al.* 1950). Detection by the counters of these emissions can reduce the counting accuracy by causing a high random to true coincidence ratio and by detecting true coincidences from the Be⁹(d,p)Be¹⁰*γBe¹⁰ reaction.

A preliminary examination of the γ-radiation in the region of interest was carried out using a fast 10 channel analyser having general characteristics similar to the analyser described by Elmore and Sands (1949).

The analysis, which was made with overlapping runs on 5 V. channel widths, is shown in Figure 2, together with the 364 keV. γ-line from I¹³¹ which was used for calibration. The γ-peaks at 480 keV. from Li⁷* and 710 keV. from B¹⁰* are clearly resolved from the background. As the γ-radiation associated with the Be⁹(d,p)Be¹⁰*γBe¹⁰ protons has an energy of 3·37 MeV., a fast single channel analyser was introduced into the γ-channel and set to the range 40–65 V. (Fig. 2). To prevent loss of true coincidences arising from the 0·2 μsec. resolving time of the analyser, coincidence circuits having 0·5 μsec. resolving time were used. With these resolving times the deuteron beam was

limited to approximately $0 \cdot 5 \mu A.$ to maintain a reasonable true to random coincidence ratio.

The electronic circuitry used in conjunction with the counters is shown in block schematic form in Figure 1. Pulses from the particle counter photo-multiplier are fed to a preamplifier and thence to a Dynatron type 1008 linear amplifier, the output of which passes to a discriminator. The discriminator output simultaneously operates a scaler and triggers a pulse shaping circuit to produce negative rectangular pulses which operate two coincidence circuits.

The γ-counter output passes to a preamplifier and thence to a fast main amplifier, the output of which is analysed by the single channel pulse amplitude analyser constructed according to the circuit by Roulston (1950). As in the case of the particle counter channel a rectangular pulse is formed to operate the coincidence circuits.
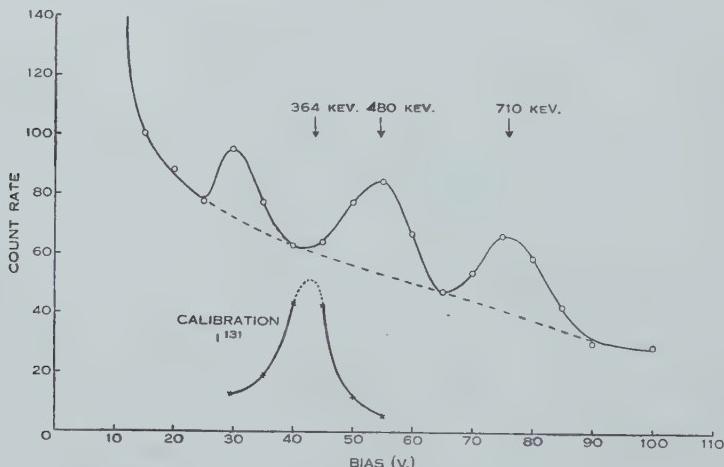
Fig. 2.—Analysis of γ-radiation from deuteron bombardment of Be⁹.

To enable an accurate allowance to be made for the random coincidence rate this quantity is measured by one coincidence circuit while another counts true plus random coincidences.

As seen in Figure 1 both particle and γ-channels are connected directly to one coincidence circuit, whilst only the γ-channel is connected directly to a second coincidence circuit, the particle channel pulse reaching the coincidence circuit in this case after travelling through a $1 \cdot 8$ μsec. delay line (approximately four times the coincidence circuit resolving time). In the first case true plus random coincidences will be counted while in the second case no true coincidences can be counted but only random ones.

III. Results and Corrections

With the target plane set at 45° to the α-particle direction, total and random coincidences were counted for several runs. Each run consisted of recording the coincidences counted with the counter set at each of several angles from 60 to 270° against a constant α-particle monitor count.

The target plane was then rotated to make an angle of 105° with the α-counter and several runs again made, each over several angles.

Because the random to total coincidence rate was high, approximately $1 : 2 \cdot 5$, an accurate knowledge of the ratio of the resolving times of the total and random coincidence circuits was desirable. This was achieved by comparing the resolving time ratio by means of random coincidences from a relatively intense source before and after the measurement at each angle. The stability of the coincidence circuits was found to be very satisfactory, the drift during a 20 minute run being in the order ± 2 per cent.

Using the random count recorded during the measurement of a total count and the mean ratio of the resolving times, a corrected random count was subtracted from the total count, giving the true count at that angle. The true counts for each run were then expressed as a fraction of the mean true count during the run. In this way errors due to variations in the gain of the Dynatron amplifier from one run to another were minimized.

This fractional expression of the true count at each angle was then corrected for eccentricity of the γ-counter with respect to the target, for γ-absorption in the target, and for centre of mass and Doppler effects.

The eccentricity at each angle was determined by measuring directly with a travelling microscope the distance between the γ-counter and the target chamber centre, and the position of the target within the chamber. The maximum correction was $2 \cdot 2 \pm 0 \cdot 1$ per cent.

The corrections for γ-absorption in the target and for change of efficiency of the γ-detector with the energy change due to the Doppler effect were calculated from the absorption coefficients given by Davisson and Evans (1952).

From the known half-life of Li⁷* (Elliott and Bell 1949) and the range of the Li⁷* recoil nucleus in beryllium, the velocity at the instant of emission of the γ-ray could be estimated and a centre of mass correction applied using the relation (from Devons and Hine 1949)

$$g(\varphi) = f\left(\varphi - \frac{v}{c} \sin \varphi\right)\left(1 - \frac{2v}{c} \cos \varphi\right),$$

where $f(\varphi)$ is the yield measured in laboratory coordinates,

  $g(\varphi)$ is the yield in centre of mass coordinates,

  v is the recoil nucleus velocity at the instant of emission,

  φ is the angle of emission of the γ-ray measured with respect to the direction of the recoil nucleus.

Similarly the Doppler energy change was calculated from

$$\Delta E = E \frac{v}{c} \cos \varphi,$$

where E is the energy of the γ-ray when emitted from a nucleus at rest,

  ΔE is the change in energy.

This energy change was converted to a γ-counter efficiency change using the absorption coefficient data mentioned above.

The maximum total of these four corrections was -5 per cent.

The corrected results are plotted against θ in Figure 3. The data represent a total of approximately 13,000 true coincidence counts. With weight factors appropriate to the number of counts at each angle the results were analysed by a least squares method to fit a curve of the form $1 + A \cos^2 \theta$. The correlation factor A so obtained was $-0 \cdot 041 \pm 0 \cdot 025$.
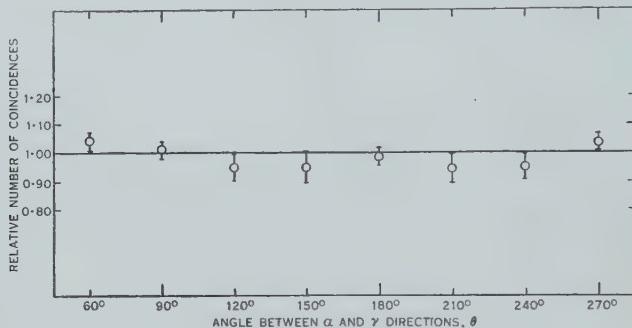
Fig. 3.—α-γ Angular correlation.

IV. DISCUSSION

The results given above show no significant departure from isotropy. While this is consistent with a Li^7 spin of $\frac{1}{2}$, one may consider circumstances which can lead to spherical symmetry if the spin is 5/2.

The anisotropic distribution of the unresolved α-particles from the deuteron bombardment of Be^9 found by Resnick and Hanna (1951) implies that α-particles and deuterons with non-zero orbital angular momentum must contribute significantly to this reaction. The plausible assumption of p wave deuterons allows three feasible values of the spin of the B^{11} compound state, only one of which could produce isotropy by chance cancellation through channel spin degeneracy.

Because of the assumption of Li^{7*} spin of 5/2 and p deuterons, it is likely that the α-particles will have orbital angular momentum $l = 1$. The angular correlation for both magnetic dipole and electric quadrupole γ-radiation will then be limited to the form $1 + A \cos^2 \theta$. Cancellation of asymmetry terms may occur by virtue of an anomalous dipole-quadrupole intensity ratio of about 25 : 1 which may be compared with the theoretical estimate of at least 200 : 1 (Inglis 1951 ; Lloyd 1951).

It follows that accidental isotropy in this reaction is improbable, which allows only the conclusion that the spin of Li^{7*} is $\frac{1}{2}$ in support of the correlation experiments discussed in the introduction.

V. ACKNOWLEDGMENTS

VI. References

BIEDENHARN, L. C., ARFKEN, G. B., and ROSE, M. E. (1951).—*Phys. Rev.* **83** : 586.

BURKE, W. H., and RISSER, J. R. (1952).—*Phys. Rev.* **87** : 294.

CLASS, C. M., and HANNA, S. S. (1951).—*Nature* **168** : 429.

CLASS, C. M., and HANNA, S. S. (1952).—*Phys. Rev.* **87** : 247.

DAVISSON, C. M., and EVANS, R. D. (1952).—*Rev. Mod. Phys.* **24** : 79.

DEVONS, S., and HINE, M. G. N. (1949).—*Proc. Roy. Soc.* A **199** : 56.

ELLIOTT, L. G., and BELL, R. E. (1949).—*Phys. Rev.* **76** : 168.

ELMORE, W. C., and SANDS, M. (1949).—" Electronics." (McGraw-Hill : New York.)

FELD, B. T. (1949).—*Phys. Rev.* **75** : 1618.

HIRST, F., and UEBERGANG, R. (1951).—*Aust. J. Sci. Res.* A **4** : 284.

HORNYAK, W. F., LAURITSEN, T., MORRISON, P., and FOWLER, W. A. (1950).—*Rev. Mod. Phys.* **22** : 291.

INGLIS, D. R. (1948).—*Phys. Rev.* **74** : 1876.

INGLIS, D. R. (1951).—*Phys. Rev.* **81** : 914.

LITTAUER, R. M. (1950).—*Proc. Phys. Soc. Lond.* A **63** : 294.

LLOYD, S. P. (1951).—*Phys. Rev.* **83** : 716.

NEWTON, J. O. (1951).—*Proc. Phys. Soc. Lond.* A **64** : 938.

PHILLIPS, G. C., HEYDENBURG, N. P., and COWIE, D. B. (1952).—*Phys. Rev.* **85** : 742.

RESNICK, I., and HANNA, S. S. (1951).—*Phys. Rev.* **82** : 463.

ROSE, B., and WILSON, A. R. W. (1950).—*Phys. Rev.* **78** : 68.

ROULSTON, K. I. (1950).—*Nucleonics* **7** (4) : 27.

THE ABSORPTION OF THE HARD COMPONENT OF COSMIC RAYS IN WATER

By A. J. Dyer*

Summary

The differential and integral range spectra of the hard component of cosmic rays in water have been measured down to a depth of 18·3 m. of water, and the integral results confirm the work of Ehmert (1937) and Wilson (1938) in this region. A lack of statistical accuracy in the differential measurements prevents the possible observation of an anomaly corresponding to 2·5 BeV./c. momentum. A comparison of the range and momentum spectra on the basis of the energy loss data of Halpern and Hall shows satisfactory agreement. The absolute value of the differential intensity is found to be 20 per cent. higher than that given by Rossi (1948) but agrees with the more recent result of York (1952).

I. Introduction

The momentum spectrum of the hard component at sea-level has been determined by magnetic deflection methods by Blackett (1937), Jones (1939), Hughes (1940), Wilson (1946), Glaser, Hamermesh, and Safonov (1950), and Caro, Parry, and Rathgeber (1951). Several authors have measured the absorption of cosmic rays in large thicknesses of various absorbers and the results of Ehmert (1937) and Wilson (1938) are generally regarded as the most reliable.

By equating analytical expressions for the momentum spectrum and the range spectrum using the results then available Janossy (1948) finds the value of $2 \cdot 0$ MeV./g.cm.$^{-2}$ at all energies for the energy loss function in water. According to Halpern and Hall (1948) the rate of energy loss of mesons in water increases from $2 \cdot 1$ MeV./g.cm.$^{-2}$ at an energy of $0 \cdot 3 \times 10^9$ eV. to $2 \cdot 6$ MeV./g.cm.$^{-2}$ at an energy of 10^{10} eV.

George (1952) has analysed the results of all measurements down to a depth of 3000 m. of water equivalent. It is shown that the soft component cannot be regarded as being in equilibrium with the hard component at very great depths, as is generally assumed (e.g. Rossi 1948). The latter assumption is based on the results of Ehmert who measured the reduction in intensity at six depths by the insertion of 5 cm. of lead absorber. George shows further that the distribution with zenith angle varies with depth, and thus all absorption experiments using wide-angle telescope geometry require correction.

Although the greatest depth of water available locally was only 20 m. it was considered profitable to carry out an absorption experiment with narrow-angle telescope geometry and using the same thickness of lead absorber (10 cm.)

---

* Physics Department, University of Melbourne.

as had been used in the magnet spectrometer measurement performed here
(Caro, Parry, and Rathgeber 1951).

It was decided also to measure the differential range spectrum to see if the
anomaly in the momentum spectrum at $2 \cdot 5$ BeV./c., referred to by several
authors (Blackett 1937 ; Glaser, Hamermesh, and Safonov 1950 ; Caro, Parry,
and Rathgeber 1951), could be detected.

II. DESCRIPTION OF EQUIPMENT

The apparatus consisted of a Geiger counter telescope which with associated
amplifiers and battery supply was enclosed in a steel tank of approximate
dimensions 6 ft. high by 2 ft. 3 in. diameter with a wall thickness of $\frac{1}{4}$ in.
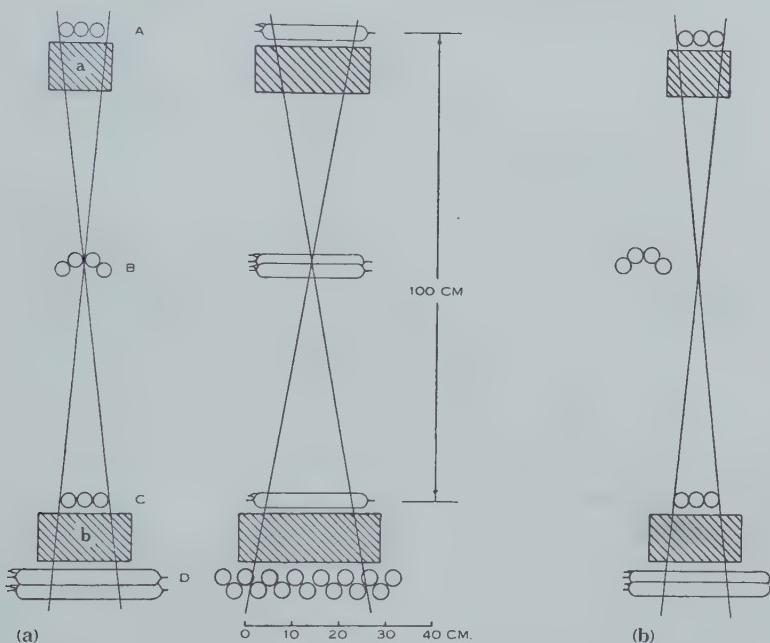
Fig. 1 (a).—Normal counter arrangement.
Fig. 1 (b).—Counter arrangement for shower measurement.

The threefold counter telescope (Fig. 1 (a)) registered particles penetrating
10 cm. of lead within an angle of 10° to the vertical, and a double layer anti-
coincidence tray was placed after a further 10 cm. of lead. Measurements
could thus be made simultaneously of the integral and differential range spectra.

The counters had a diameter of $3 \cdot 4$ cm. After they were filled with an
argon-ethylene mixture in the proportion 10 : 1, the pressure was adjusted to
give a starting voltage of 1000 ± 10 V. The operating voltage was 1150 V.

The counters in trays A, B, and C had an effective length of 18 cm. and
those in tray D an effective length of 22 cm. The lead absorber a was placed
as high in the telescope as possible to prevent scattering in this absorber affecting
the anti-coincidence count.

A preliminary experiment showed that the dimensions of tray D had to be twice those given by a linear extrapolation of the telescope geometry before the effect of scattering in absorber b could be regarded as negligible.   This is consistent with the experience of Trumpy and Orlin (1943) and the dimensions of tray D in the experimental set-up were arranged to satisfy this condition.

Conventional Rossi circuits, each with a resolving time of 10 μsec., recorded threefold coincidences between trays A, B, and C, and between these and tray D.

A photograph of the mechanical recorders was taken every hour on " Super-XX " 16-mm. film by fitting an accurate chronometer with contacts which actuated the lights and camera through a relay system.
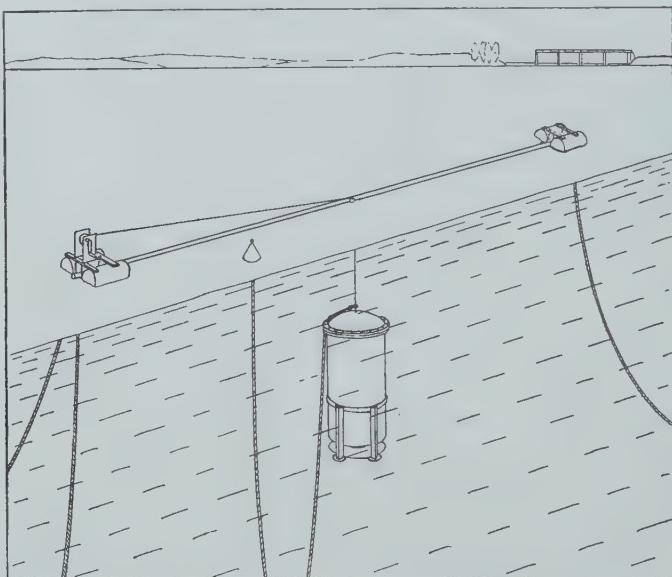
Fig. 2.—Arrangement for submerging apparatus.

The equipment was submerged at various depths down to a maximum of 18·3 m. in a freshwater reservoir at Melton, Victoria, the greatest depth being obtained during the winter of 1951.   The apparatus, which had a weight of about 30 lb. when submerged, was supported by a $\frac{1}{4}$ in. steel wire from a light wooden beam $30\frac{1}{2}$ ft. long and was raised and lowered by a hand winch mounted at one end of the boom.

A safety rope (Fig. 2) with a buoy was attached to the tank and this had to be used on three occasions when the buoyancy tanks supporting the boom were punctured.   The boom was provided with two anchors upstream and one downstream.   A pontoon-type surface craft fitted with an overhead hoist could be brought over the boom and the apparatus hauled on deck for maintenance.

III. Experimental Results

Measurements were made of the integral and differential range spectra, up to an absorber thickness of 1900 g.cm.$^{-2}$ of water.   The background counting rates for both measurements were taken as being the corresponding counting

rates when the coincidence trays were placed out of line (Fig. 1 (b)). The significance of this method is discussed in Section IV.

An additional background effect in the differential measurement was allowed for by taking a series of measurements with absorber b removed. This background was $2 \cdot 40 \pm 0 \cdot 05$ counts/hr. at sea-level and was approximately proportional to the integral counting rate at other depths.

CURVE DEDUCED FROM MOMENTUM SPECTRUM
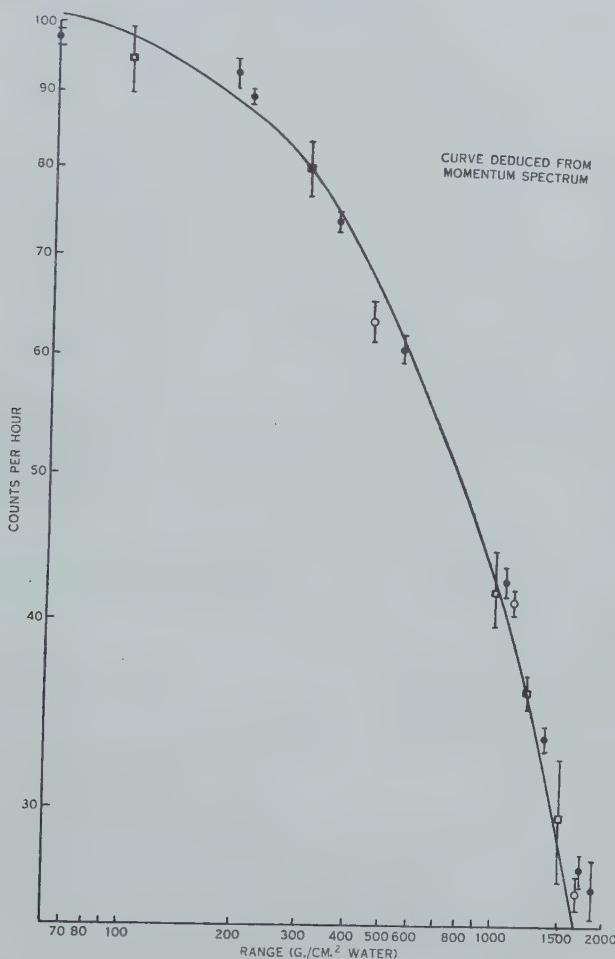
Fig. 3.—Integral range spectrum.
● Author. □ Ehmert. ○ Wilson.

The water equivalent absorber of the 10-cm. lead absorber was calculated to be 60 g.cm.$^{-2}$ from the data given by Rossi (1948) and Halpern and Hall (1948).

The final results are plotted in Figures 3 and 4 and presented in Tables 1 and 2.

In Figure 3 the integral results are compared with those of Ehmert (1937), Wilson (1938), and a curve deduced from the momentum spectrum determination

E
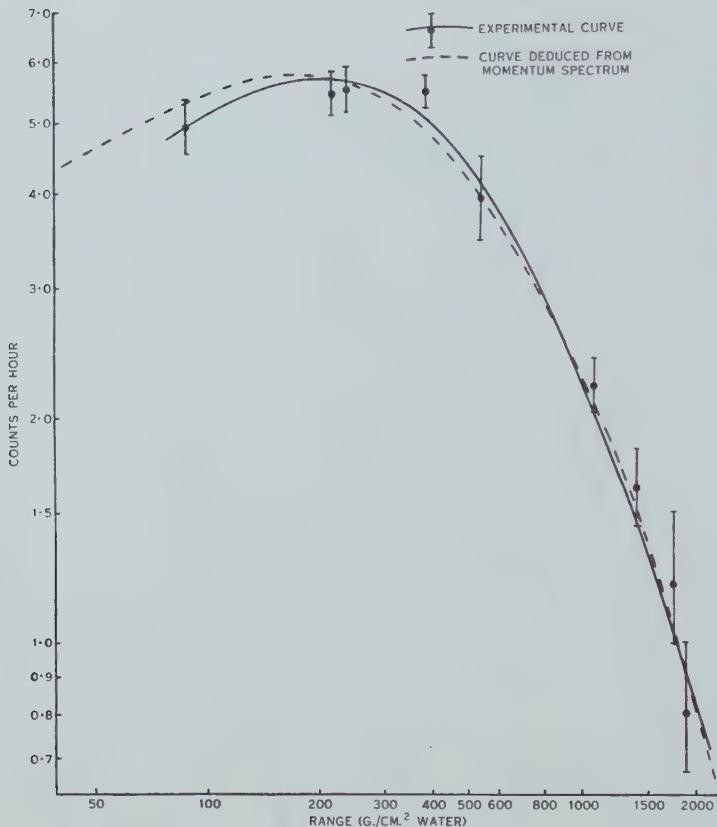
Fig. 4.—Differential range spectrum.

of Caro, Parry, and Rathgeber (1951). In order to convert the latter results into range spectra, the range energy relation for mesons in water was set up from the curves given by Halpern and Hall (1948).

TABLE 1

BACKGROUND MEASUREMENTS (COUNTS/HR.)

Depth (m. of water)	Showers		Absorber b Removed. Differential	Total Differential Background
	Integral	Differential		
0	$11 \cdot 1 \pm 0 \cdot 5$	$0 \cdot 71 \pm 0 \cdot 08$	$2 \cdot 40 \pm 0 \cdot 05$	$3 \cdot 11 \pm 0 \cdot 10$
2·1	$8 \cdot 8 \pm 0 \cdot 4$	$0 \cdot 65 \pm 0 \cdot 08$	$1 \cdot 93 \pm 0 \cdot 04$	$2 \cdot 58 \pm 0 \cdot 10$
3·1	$7 \cdot 8 \pm 0 \cdot 4$	$0 \cdot 60 \pm 0 \cdot 08$	$1 \cdot 77 \pm 0 \cdot 04$	$2 \cdot 37 \pm 0 \cdot 09$
6·1	$5 \cdot 6 \pm 0 \cdot 2$	$0 \cdot 34 \pm 0 \cdot 06$	$1 \cdot 25 \pm 0 \cdot 03$	$1 \cdot 59 \pm 0 \cdot 07$
9·1	$4 \cdot 9 \pm 0 \cdot 3$	$0 \cdot 28 \pm 0 \cdot 05$	$1 \cdot 10 \pm 0 \cdot 02$	$1 \cdot 38 \pm 0 \cdot 06$
12·2	$4 \cdot 1 \pm 0 \cdot 3$	$0 \cdot 25 \pm 0 \cdot 05$	$0 \cdot 88 \pm 0 \cdot 02$	$1 \cdot 13 \pm 0 \cdot 06$
15·3	$3 \cdot 8 \pm 0 \cdot 3$	$0 \cdot 20 \pm 0 \cdot 05$	$0 \cdot 73 \pm 0 \cdot 02$	$0 \cdot 93 \pm 0 \cdot 06$

TABLE 2

INTEGRAL AND DIFFERENTIAL RANGE SPECTRA

Depth (m. of water)	Integral		Differential	
	(Counts/hr.)	Range (g.cm.$^{-2}$ of water)	(Counts/hr.)	Range (g.cm.$^{-2}$ of water)
0·0	97·8±0·6	60	4·9±0·4	90
1·3	93·2±2·0	190	5·4±0·4	220
1·5	89·7±1·0	210	5·5±0·4	240
3·0	74·8±0·8	360	5·4±0·2	390
4·9	60·9±1·0	550	3·9±0·6	580
10·0	42·5±1·0	1060	2·2±0·3	1090
13·4	33·5±1·0	1400	1·6±0·3	1430
16·8	27·4±1·0	1740	1·2±0·3	1770
18·3	26·6±2·0	1890	0·8±0·2	1920

The differential results are presented in Figure 4, together with a comparison curve from the momentum spectrum (dotted line).

IV. Discussion of Results

(a) Integral Spectrum

It can be seen from Figure 3 that the results are in satisfactory agreement with those of Ehmert (1937) and Wilson (1938), whose measurements were made without any lead absorber. The soft component may therefore be regarded as being in equilibrium with the hard component in this region, confirming the reports of Rossi (1948) and George (1952). The agreement between the present results and the curve deduced from the momentum spectrum is also satisfactory.

The shower measurements recorded here and used as a background correction were taken with the counter trays out of line in an attempt to avoid under-estimating the penetrating shower component. This was done since the magnet spectrometer possesses a stronger bias against these events than the simple telescope geometry, and the absorption of showers is known to take place less rapidly than the single particle intensity (George 1952).

As the penetrating shower rate is found to be roughly proportional to the single particle rate at small depths of water, the shape of the absorption curve is not significantly affected here. The relative shower intensity observed was, however, about five times greater than the similar measurement without lead absorber by Ehmert who used a much smaller shower selection geometry, and would tend to remove the discrepancy found by Rathgeber (1951) who compared the absorption and momentum spectra at greater depths using the data of Halpern and Hall.

No such discrepancy is found by George (1952) but his values for the energy loss were taken from Fermi and do not exhibit the so-called logarithmic increase

of Halpern and Hall which appears to be confirmed by recent experiments (Becker *et al.* 1952 ; Taylor 1952).

(b) *Differential Spectrum*

The differential range spectrum is consistent with the momentum spectrum, but the results do not possess sufficient statistical accuracy for the observation of any anomaly at $2 \cdot 5$ BeV./c. momentum (approx. 1060 g.cm.$^{-2}$). The equipment is at present being modified for further examination of this feature.

By making use of the integral results to normalize to the data of Rossi (1948), the absolute differential intensity at 100 g.cm.$^{-2}$ of air equivalent is found to be $6 \cdot 55$ g.$^{-1}$sec.$^{-1}$sterad.$^{-1}$ in terms of air equivalent, which is 20 per cent. higher than Rossi's value of $5 \cdot 40$ g.$^{-1}$sec.$^{-1}$sterad.$^{-1}$. A recent measurement of the differential range spectrum in lead by York (1952) similarly gives a result 20 per cent. greater than the Rossi data. York concludes that the low value given by Rossi is due to the fact that no correction was made for particles scattered out of the counter telescope in the experiment on which his value is based. This opinion is supported by the present measurement in which considerable care was taken to ensure that only a negligible proportion of particles were scattered beyond the coverage of the anti-coincidence tray.

V. Acknowledgments

VI. References

Becker, M. J., Chanson, M. P., Nageotte, M. E., Treille, M. P., Price, B. T., and Rothwell, P. (1952).—*Proc. Phys. Soc. Lond.* A **65** : 437.

Blackett, P. M. S. (1937).—*Proc. Roy. Soc.* A **159** : 2.

Caro, D. E., Parry, J. K., and Rathgeber, H. D. (1951).—*Aust. J. Sci. Res.* A **4** : 16.

Ehmert, A. (1937).—*Z. Phys.* **106** : 751.

George, E. P. (1952).—" Progress in Cosmic Ray Physics." (Ed. J. G. Wilson.) (Interscience Publishers : New York.)

Glaser, D. A., Hamermesh, B., and Safonov, G. (1950).—*Phys. Rev.* **80** : 625.

Halpern, O., and Hall, H. (1948).—*Phys. Rev.* **73** : 477.

Hughes, D. J. (1940).—*Phys. Rev.* **57** : 592.

Janossy, L. (1948).—" Cosmic Rays." (Oxford Univ. Press.)

Jones, H. (1939).—*Rev. Mod. Phys.* **11** : 235.

Rathgeber, H. D. (1951).—*Z. Naturf.* **6a** : 599.

Rossi, B. (1948).—*Rev. Mod. Phys.* **20** : 537.

Taylor, A. E. (1952).—*Rep. Progr. Phys.* **15** : 49.

Trumpy, B., and Orlin, J. (1943).—*Bergens Mus. Aarb. Naturvidensk.* **1943**, No. 7.

Wilson, J. G. (1946).—*Nature* **158** : 414.

Wilson, V. C. (1938).—*Phys. Rev.* **53** : 337.

York, C. M. (1952).—*Phys. Rev.* **85** : 998.

AN ESTIMATE OF THE DENSITY AND MOTION OF SOLAR MATERIAL FROM OBSERVED CHARACTERISTICS OF SOLAR RADIO OUTBURSTS

By Hari K. Sen*

Summary

The theory of radio wave generation by multistream charge interaction (Feinstein and Sen 1951) is extended and applied to the observations made by Australian workers (Wild 1950) of the spectrum of outbursts of solar radio-frequency radiation in the frequency range 70–130 Mc/s. The dispersion equation is derived as a function of the velocity of solar material erupting into a static corona and of the temperatures and densities of the material and the corona. The application of the dispersion equation to the Australian data (loc. cit.) enables an estimate to be made of the velocity ($\simeq$500 km./s.) and the particle density ($\simeq 10^8$ cm.$^{-3}$) of the moving solar material.

I. Introduction

The high thermal temperature (10^8–10^{13} °K.) associated with solar noise bursts and outbursts has led several workers to suggest a non-equilibrium mechanism for their origin (Pawsey 1950). The principal one in this field has been space-charge wave amplification in moving interacting beams as in the electron-wave tube (Haeff 1949a). Recently, Australian workers (Wild 1950) have obtained, by a sweep-frequency technique, the spectrum of " outbursts " of solar radio-frequency radiation in the frequency range 70–130 Mc/s. They define an " outburst " as a " burst having a particular type of ' dynamic ' spectrum, characterized by a drift of spectral features, with time, towards the lower frequencies at a rate of the order of $\frac{1}{4}$ Mc/s. per second ". They tentatively interpret the spectra in terms of the accelerated motion of particles that finally leave the solar atmosphere and cause terrestrial magnetic storms.

The author in the present paper makes an attempt to interpret the Australian data as a phenomenon of radio wave generation by solar material moving through the ionized plasma of the corona. He believes that such an interpretation provides an independent method through radio observations of estimating the density and the motion of the solar corpuscles *in the solar atmosphere*, that, later on, are supposed to cause terrestrial magnetic storms. Theories of the latter, it is known, give us estimates of the motion of the corpuscles in transit from the Sun to the Earth and of their density near the Earth (Chapman and Bartels 1940 ; Kiepenheuer 1952). The author finds that his solar estimates are consistent with the terrestrial ones.

* National Bureau of Standards, Washington 25, D.C., U.S.A.

II. Theory

We shall first derive the appropriate dispersion equation* as a function of the velocity of the moving (ionized but macroscopically neutral) solar material and the temperatures and densities of the material and the corona. This dispersion equation we shall consider as a particular case of a more general one, that is, that of n interacting beams of electrons. We shall neglect the motion of the ions, on account of their large mass,† and assume that there is no static space charge, the charge on the ions balancing that on the electrons.

Let the physical quantities, viz. charge density, velocity, and current density of the nth interacting beam be denoted, respectively, by

$$\left.\begin{array}{l} \rho'_n = \rho_n + \bar{\rho}_n, \quad v'_n = v_n + \bar{v}_n, \\[2mm] i'_n = i_n + \bar{i}_n = i_n + \bar{\rho}_n v_n + \rho_n \bar{v}_n, \end{array}\right\} \quad \dots\dots\dots\dots (1)$$

where the bars denote the A.C. perturbations. We shall suppose these perturbations to be of the first order of small quantities, and neglect all quantities of order higher than the first.

Assume that all A.C. quantities vary as

$$e^{-\Gamma z + j\omega t} \quad , \quad \dots\dots\dots\dots\dots (2)$$

where the z-axis is the direction of propagation, Γ the propagation constant, and ω the angular frequency of the disturbance.

In the absence of a static charge, Poisson's equation gives (in rationalized m.k.s. units)

$$\varepsilon_0 \frac{\partial \bar{E}}{\partial z} = \sum_1^n \bar{\rho}_n, \quad \dots\dots\dots\dots\dots (3)$$

where $\bar{E}$ is the perturbed electric field and ε_0 the dielectric constant of free space.

The charge distribution of the nth beam must satisfy the equation of continuity:

$$\frac{\partial \rho'_n}{\partial t} + \frac{\partial i'_n}{\partial z} = 0. \quad \dots\dots\dots\dots\dots (4)$$

From (1), (2), and (4), we derive

$$(j\omega - \Gamma v_n)\bar{\rho}_n - \Gamma \bar{v}_n \rho_n = 0. \quad \dots\dots\dots\dots (5)$$

We assume the plasma to be of sufficiently low density so that we can neglect the frictional forces. Then we have the equation of motion for the nth beam:

$$\frac{\partial v'_n}{\partial t} + v'_n \frac{\partial v'_n}{\partial z} + \frac{1}{\rho'_n} \frac{\partial}{\partial z}\left(\frac{kT_n}{m}\rho'_n\right) = \frac{e}{m} E', \quad \dots\dots (6)‡$$

---

* The physical theory of space-charge wave amplification in moving interacting charges has been well discussed in the literature. The author is fully aware of the differences of opinion in the field, particularly with reference to the application of laboratory data to solar atmospheres. Discussion of such niceties is beyond the scope of the present paper.

† The ions contribute terms that are $(\omega_i/\omega_e)^2$ times the electronic terms in equation (10) of this paper, where ω_i is the ionic and ω_e the electronic plasma frequency given by equation (11).

‡ Equation (6) results on multiplying the Boltzmann equation:

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla f + \frac{e}{m}\mathbf{E} \cdot \frac{\partial f}{\partial \mathbf{v}} = 0$$

by $\mathbf{v}$ and integrating with respect to $\mathbf{v}$.

where T_n is the temperature of the nth beam, e^* the charge and m the mass of the electron, and k is the Boltzmann constant.

Substitute (1) in (6) and obtain

$$\frac{\partial \bar{v}_n}{\partial t} + v_n \frac{\partial \bar{v}_n}{\partial z} + \frac{1}{\rho_n} \frac{\partial}{\partial z} \left(\frac{kT_n}{m} \bar{\rho}_n \right) = \frac{e}{m} \bar{E}, \quad \dots \dots \dots \dots (7)$$

as the static space charge is assumed to be zero.

For the variation (2) of the A.C. quantities, equation (7) reduces to

$$(j\omega - \Gamma v_n)\bar{v}_n - \frac{\Gamma}{\rho_n} \frac{kT_n}{m} \bar{\rho}_n = \frac{e}{m} \bar{E}. \quad \dots \dots \dots \dots (8)$$

Eliminate $\bar{v}_n$ between (5) and (8) and obtain for the A.C. charge density :

$$\bar{\rho}_n = \frac{(e/m)\bar{E}\rho_n}{(j\omega - \Gamma v_n)^2/\Gamma - \Gamma(kT_n/m)}. \quad \dots \dots \dots \dots (9)$$

From (3) and (9), we have the dispersion equation :

$$\sum_1^n \frac{\omega_n^2}{(\omega + j\Gamma v_n)^2 + (kT_n/m)\Gamma^2} = 1, \quad \dots \dots \dots \dots (10)$$

where ω_n is the plasma angular frequency, given by

$$\omega_n^2 = \frac{e}{m\varepsilon_0}\rho_n. \quad \dots \dots \dots \dots \dots (11)$$

For solar material of density ρ_2 and temperature T_2, erupting with velocity v into a static corona of density ρ_1 and temperature T_1, the dispersion equation (10) reduces to

$$\frac{\omega_1^2}{\omega^2 + (kT_1/m)\Gamma^2} + \frac{\omega_2^2}{(\omega + j\Gamma v)^2 + (kT_2/m)\Gamma^2} = 1. \quad \dots \dots (12)$$

Introduce in (12) the following dimensionless quantities :

$$\left. \begin{array}{ll} \gamma = \dfrac{j\Gamma v}{\omega}, & \alpha_n = \dfrac{kT_n}{mv^2}, \\[2ex] \beta_n = \dfrac{\omega^2}{\omega_n^2}, & x = \dfrac{\beta_2}{\beta_1}. \end{array} \right\} \quad \dots \dots \dots \dots (13)$$

The equation (12) then reduces to a quartic in γ :

$$\alpha_1(\alpha_2 - 1)\beta_1 x \gamma^4 - 2\alpha_1 \beta_1 x \gamma^3 - \{\alpha_1(\beta_1 x - 1) + (\alpha_2 - 1)(\beta_1 - 1)x\}\gamma^2$$
$$+ 2(\beta_1 - 1)x\gamma + (\beta_1 - 1)x - 1 = 0. \quad \dots \dots \dots \dots (14)$$

Equation (14) will give the range of β_1 for which γ is complex for a fixed set of values of the parameters α_1, α_2, and x.

III. RESULTS AND CONCLUSIONS

We assume T_1 (electron temperature of corona) $\simeq 10^6$ °K.,[†] and T_2 (temperature of moving material) $\simeq 10^4$ °K. To fit the Australian data we have to determine our disposable parameters v and x, which will give an estimate of the motion and the density of the solar corpuscles. Table 1 gives a summary of the results obtained.

---

* With the proper sign (— in this case).

† We have neglected the small variation in coronal electron temperature with height.

The second column of Table 1 gives the electron density, n_1, of the corona as computed from the Allen-Baumbach formula (Allen 1947) :

$$n_1 = 10^8(1 \cdot 55\rho^{-6} + 2 \cdot 99\rho^{-16}) \text{ electrons cm.}^{-3}, \quad \ldots \ldots (15)$$

where ρ is the distance from the centre of the Sun in units of the solar radius. The fourth column gives the electron concentration, n_2, of the moving material at the different heights. The density was fixed at an arbitrary height over the solar surface ($1 \cdot 68 \times 10^8$ cm.$^{-3}$ at $11 \cdot 6 \times 10^4$ km.), and the densities at the remaining heights were calculated according to the inverse square law dilution from the solar centre. The fixing of the density was done by trial and error so as to get amplification within the observed range of frequencies. The sixth column gives the velocities, v, of the moving material at the different heights

TABLE 1

1	2	3	4	5	6	7	8
Height above Sun's Surface h (km.)	Electron Density of Corona n_1 (cm.$^{-3}$)	Plasma Frequency f_1 (Mc/s.)	Electron Density of Moving Material n_2 (cm.$^{-3}$)	$x = \dfrac{n_1}{n_2}$	Velocity of Moving Material v (km./s.)	Velocity of Ca$^+$ (Milne's Formula) (km./s.)	Frequency Bandwidth of Amplification f (Mc/s.)
$\times 10^4$	$\times 10^8$		$\times 10^8$				
$8 \cdot 0$	$1 \cdot 34$	$104 \cdot 0$	$1 \cdot 83$	$0 \cdot 73$	305		$93 \cdot 3 – 159 \cdot 7$
$10 \cdot 0$	$1 \cdot 03$	$91 \cdot 0$	$1 \cdot 74$	$0 \cdot 59$	388	86	$86 \cdot 2 – 147 \cdot 7$
$11 \cdot 6$	$0 \cdot 84$	$82 \cdot 5$	$1 \cdot 68$	$0 \cdot 50$	500	244	$80 \cdot 5 – 141 \cdot 3$
$14 \cdot 0$	$0 \cdot 67$	$73 \cdot 0$	$1 \cdot 58$	$0 \cdot 42$	600	366	$74 \cdot 6 – 131 \cdot 8$
$16 \cdot 0$	$0 \cdot 55$	$66 \cdot 5$	$1 \cdot 51$	$0 \cdot 36$	740	439	$69 \cdot 3 – 125 \cdot 4$

that conform with the drift of the cut-off frequency towards lower frequencies at the rate of $0 \cdot 22$ Mc/s. per second (Wild 1950, p. 402). The cut-off frequency is supposed to be the plasma frequency, f_1, at the level concerned (third column).* For comparison, the velocities have been given in the seventh column as calculated from Milne's (1926) theory of the expulsion of calcium ions by radiation pressure. The eighth column gives the frequency range, f ($= \omega/2\pi$), within which amplification is possible.

Figure 1 gives the amplification v. frequency curves for the different heights and velocities listed in Table 1. The ordinate is the amplification factor $A = (\omega/v)\gamma_2$, where γ_2 is the imaginary part of γ in (14). With increasing velocity, the curves shift toward the lower frequencies. This is in conformity with the observed drift of spectral features, with time, towards the lower frequencies, and the tendency of the spectrum to show signs of subsiding at the higher frequencies (Wild 1950). Also, observational evidence on prominence motion and the theory of magnetic storms indicate an outward acceleration of solar material.

* $f_1 = \omega_1/2\pi$, where ω_1 is given by (11).

The velocities (305–740 km./s.) and the electron density ($\simeq 10^8$ cm.$^{-3}$) of the moving material *in the solar atmosphere* should be compared with the velocity (1000–1500 km./s.) and density (100 cm.$^{-3}$) of solar corpuscles near the Earth, as estimated from the theory of magnetic storms (Chapman and Bartels 1940 ; Kiepenheuer 1952). Whipple and Gossner (1949) obtain, from the scattering of sunlight by electrons, the upper limit of electron density near the Earth $\simeq 10^3$ cm.$^{-3}$, which is in close accord with our value.

No particular significance need be attached to the values of the amplification factors. A first-order theory as developed above can only indicate the qualitative trend. We shall need a non-linear theory to achieve quantitative accuracy. Nevertheless, we can estimate if it is possible to receive the observed radio flux on the Earth, on the assumption that the source of the available energy is the initial difference of energy between the interacting streams (Haeff 1949*b*). Let
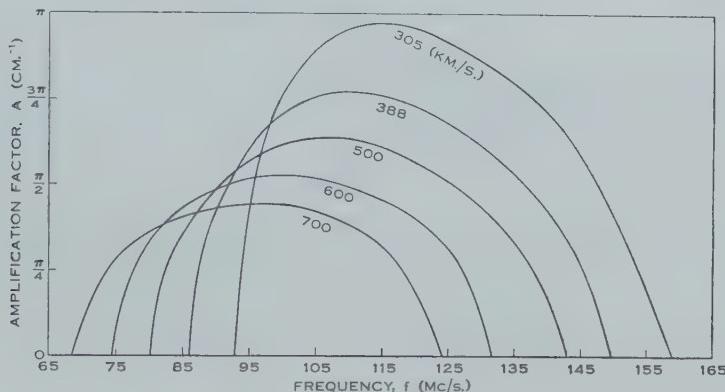
Fig. 1.—Curves of amplification A v. frequency f, for different velocities of solar material ; wave growth $= \exp[Az]$.

us assume an efficiency factor α for this conversion. Then the power, P, generated per unit bandwidth at solar surface by solar material erupting into a static corona, is given by

$$P = \tfrac{1}{2}\alpha N m v^3 S/\Delta\nu, \quad \dots\dots\dots\dots\dots\dots (16)$$

where N is the density, m the mass, and v the velocity of the electrons, $\Delta\nu$ is the frequency bandwidth within which the power is radiated, and S is the area of the active surface ejecting the corpuscles.

Hence, if R is the Earth-Sun distance, the power flux at the Earth is

$$F = \frac{P}{4\pi R^2} = \tfrac{1}{2}\alpha N m v^3 \frac{S}{4\pi R^2 \Delta\nu}. \quad \dots\dots\dots\dots\dots (17)$$

Setting $N = 10^8$ cm.$^{-3}$, $m = 9 \cdot 12 \times 10^{-28}$ gm., $v = 500$ km./s., $S =$ one hundred-millionth of solar surface $= \dfrac{4\pi(0 \cdot 695 \times 10^{11})^2}{10^4}$ cm.^{2}, $R = 149 \cdot 5 \times 10^6$ km., and $\Delta\nu = 60$ Mc/s. in equation (17), we have

$$F = 2 \times 10^{-16} \alpha \text{ W.m.}^{-2} \text{ (c/s.)}^{-1}. \quad \dots\dots\dots\dots (18)$$

From the Australian data:

$$F \simeq 10^{-19} \text{ to } 10^{-20} \text{ W.m.}^{-2} \text{ (c/s.)}^{-1}. \quad \ldots\ldots\ldots\ldots \quad (19)$$

Hence the range of α is

$$\alpha \simeq 5 \times 10^{-4} \text{ to } 5 \times 10^{-5}. \quad \ldots\ldots\ldots\ldots\ldots \quad (20)$$

We see that enough kinetic energy is available in the moving beam to account for the observed radio flux at the Earth.

We may finally remark that during an outburst particles may be thrown out from the active area in spurts having different concentrations and velocities. This may result in several " drifting " peaks as observed by the Australians. Further, in the course of the computations, it was found that, for a fixed velocity, the frequency spectrum shifted towards higher frequencies for higher particle concentrations. This indicates the possibility of outbursts even in the centimetric range for very high particle concentrations.

IV. Acknowledgments

V. References

Allen, C. W. (1947).—*Mon. Not. R. Astr. Soc.* **107** : 426.

Chapman, S., and Bartels, J. (1940).—" Geomagnetism." Vol. 2, Ch. 25. (Oxford Univ. Press.)

Feinstein, J., and Sen, H. K. (1951).—*Phys. Rev.* **83** : 405.

Haeff, A. V. (1949a).—*Proc. Inst. Radio Engrs.* N.Y. **37** : 4.

Haeff, A. V. (1949b).—*Phys. Rev.* **75** : 1547.

Kiepenheuer, K. O. (1952).—*J. Geophys. Res.* **57** : 113.

Milne, E. A. (1926).—*Mon. Not. R. Astr. Soc.* **86** : 459.

Pawsey, J. L. (1950).—*Proc. Instn. Elect. Engrs.* **97** (III) : 305.

Whipple, F. L., and Gossner, J. L. (1949).—*Astrophys. J.* **109** : 380.

Wild, J. P. (1950).—*Aust. J. Sci. Res.* A **3** : 399.

RADAR OBSERVATIONS OF RAIN FROM NON–FREEZING CLOUDS

By R. S. STYLES* and F. W. CAMPBELL*

Summary

It is now conceded that drizzle or light rain may fall from clouds which do not reach freezing level, but it is still not generally accepted that heavy precipitation is also possible from such clouds. This paper adds five further airborne radar observations of non-freezing rain (in the vicinity of Sydney, Australia ; latitude 35 °S.) to the relatively few accounts of moderate to heavy precipitation already recorded as falling from such clouds.

Observations described are of rain, which at its peak reached " very excessive " proportions at ground level, falling from clouds, 9000–10,000 ft. deep, which had developed in a maritime air mass either over sea or slightly inland. The shape of the radar echo observed from the rain in these clouds was that of a vertical column. The radar echo reached its maximum development at or just prior to the time of maximum development of the cloud, following which both the cloud and the radar echo immediately commenced to subside or dissipate. Maximum precipitation rate at ground level generally occurred after the cloud had passed its maximum development. While the storm was developing the radar echo was generally characterized by a maximum in the middle levels of the cloud, but during the decay stage the radar echo became more uniform with height.

These observations support the theory that the mechanism of rain formation at work in such clouds is that of coalescence of water drops.

I. INTRODUCTION

Over the last decade or so, a number of accounts have appeared of moderate to heavy rain falling from clouds which were wholly warmer than freezing (Heywood 1940 ; Simpson 1941 ; Kotsch 1947 ; Hunt 1949 ; Bowen 1950 ; Smith 1951 ; Ward 1952). With a few exceptions, these accounts have been brief statements of fact, that rain of certain intensity had fallen from non-freezing clouds. Bowen (1950) showed that coalescence of water drops was the mechanism at work in the formation of such rain, supporting his statement with calculations of raindrop growth under typical cloud conditions and with four examples of non-freezing rain observed with the aid of ground and airborne radar. Later, Smith (1951) described in detail three other aerial observations.

The purpose of the present paper is to document five further observations of rain falling from clouds that were kept under observation during the major

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney. Soon after the submission of this paper it was learnt with deep regret that Mr. Styles and Mr. Campbell lost their lives in the course of further experimental work similar to that described in this paper. Their aircraft crashed into the sea some 15 miles south of Cronulla, N.S.W., on the morning of October 27, 1952, soon after entering heavy cloud on which they were making measurements. With them were the crew of the aircraft, who had undertaken many flights of a similar character : Flight-Lieutenant A. Tafe (Capt.), Squadron Leader P. Fisher, Warrant Officer W. Bocquet, and Aircraftwoman Marie Costello.

part of their life from an aircraft fitted with radar. At no time did any of these clouds reach the 0 °C. isotherm. An attempt is made to describe in as much detail as possible the important features of each rain storm and to summarize those features which were common to all five storms.

II. METHODS OF OBSERVATION

The observations described in this paper were carried out with an airborne radar equipment. An aircraft installation has great advantages for the investigation of the physics of cloud and rain formation, as the aircraft may search out the most interesting clouds and fly over and through them to observe directly their external and internal conditions. The radar is an SCR717 set operating on a wavelength of $9 \cdot 1$ cm. Its antenna is mounted on the nose of the aircraft, so that the aerial beam rotates about an axis coinciding with the longitudinal axis of the aircraft. Thus, when the aircraft is flying in a horizontal attitude, the aerial beam scans through a vertical section of the atmosphere at right angles to the line of flight. Any radar echo from rain lying in this vertical sector at any instant, be it either above, below, or on either side of the aircraft, is displayed on a Vertical Position Indicator oscilloscope in the same position relative to the centre of the oscilloscope as the actual rain area bears to the aircraft. The salient features of this type of display are shown in Plate 1. Plate 1 (a) represents a vertical section of the atmosphere and shows the aircraft fitted with the radar set flying over a raining cloud during a typical observational flight. Plate 1 (b) shows a photograph of the corresponding display on the Vertical Position Indicator oscilloscope. The aircraft appears as a spot at the centre of the oscilloscope. Owing to finite side lobes of the aerial beam, the echo from the ground appears not as a horizontal band but as a circle around the aircraft of radius equal to the height of the aircraft above the ground. The echo from the rain in and falling from the cloud is shown extending from just below the aircraft down to the ground.

The intensity of radio waves back-scattered from raindrops is proportional to $\Sigma N D^6$, where N is the number of drops per unit volume and D is their diameter. Hence raindrops would first be detected in a cloud when $\Sigma N D^6$ is sufficient to give a measurable signal.

The sensitivity of the radar used in these observations is such that raindrops of 1 mm. diameter would just be detected at a range of 10,000 ft. if they completely filled the radar beam and if the drop density were 100 drops/m.³ and drops of $0 \cdot 5$ mm. diameter would just be visible if their density at the same range was 10^4 drops/m.³

The Vertical Position Indicator display shown in Plate 1 (b) is not suitable for accurate measurement of the echo intensity at any particular height in the rain shower. To facilitate such measurements, an auxiliary oscilloscope (Plate 1 (c)) is installed on which is displayed the amplitude of the echo at various ranges from aircraft.

From this display, the sensitivity of which is known, may be calculated an equivalent echo power v. range curve, allowance being made for the known

variation of echo intensity with distance from the point of observation. For the purposes of this paper all echo power figures are reduced to an arbitrary range of 10,000 ft. In this way, the range effect is eliminated and hence a curve of equivalent echo power *v.* height above mean sea-level may be constructed. Such a curve clearly depicts the variation of the radar echo intensity down through the cloud. In the observations described in this paper, this oscilloscope (Plate 1 (*c*)) operates over a 90° sector of the plane scanned through by the aerial beam. The angular position of this 90° sector can be varied at will thus enabling the radar operator to aim it at the particular rain echo under investigation. The position of this sector is identified on the Vertical Position Indicator oscilloscope by a bright radial line called the " strobe marker " (see Plate 1 (*b*)). For convenience this strobe marker is positioned 180° away from the true centre of the 90° sector.

The two oscilloscopes described are mounted on a panel together with instruments which give the time, the altitude heading and airspeed of the aircraft, and the ambient temperature. This recorder panel may be photographed at intervals of 3, 6, 15, or 30 sec. as desired, to obtain a permanent record for subsequent analysis.

The estimates of raindrop size given in this paper were made by observing the size of the splashes on the windscreen of the aircraft when flying through the rain. Although such estimates depend entirely on the judgement of the observer, test flights near the ground have established that this technique is satisfactory for an approximate classification of drop sizes into size ranges of about 0·25 mm., 0·25–1 mm., and greater than 1 mm. The impact of drops on the aircraft skin becomes audible for drops about 1 mm. in diameter.

The calculations of the precipitation rate at ground level were based on the intensity of the radar echo at this level assuming that the raindrop spectrum was that given by Laws and Parsons (1943). Twomey (1952) has since shown that such estimates of rainfall intensity cannot be made to better than a factor of 2 : 1, owing to short period fluctuations in the actual drop spectrum on the ground.

The temperature structure of the atmosphere was obtained by a temperature sounding taken during the aircraft ascent. The thermometer used was a calibrated aircraft resistance thermometer accurate to 1 °C. No readings were taken in cloud because an aircraft thermometer is known to be unreliable when wet. This sounding is supplemented by the radiosonde data obtained at Rathmines which is situated on the coast 70 miles north of Sydney. For ready comparison of the two soundings, radiosonde data were reduced to equivalent heights in feet above mean sea-level assuming an I.C.A.N. atmosphere.

III. Five Observations of Rain from Non-freezing Clouds
(a) *January 11, 1951*

A gentle maritime stream, associated with an anticyclone centred in the southern Tasman sea, was moving across the Sydney coast from a south-easterly direction. At 0900 hours, showers were falling from 8/8 strato-cumuli and column-type rain echoes extending from the ground to approximately 8000 ft.

were observed on a 10-cm. ground radar set. A flight was commenced at 1030 hours, at which time the layer of strato-cumuli over the coast was broken but still 8/8 about 10 miles inland. The height of the cloud base was 1200 ft. with tops varying from 4000 to 11,000 ft. over the coast. Over the land the 8/8 cloud layer was generally held down at 10,000 ft. by a temperature inversion, with an occasional head rising to 11,000 ft. where the temperature was $+5\frac{1}{2}$ °C. The temperature sounding obtained during the aircraft ascent is shown in Figure 1 (a) superimposed on the sounding obtained from the 1800-hours radiosonde release at Rathmines. The radiosonde sounding indicated that freezing level was at 13,800 ft. (I.C.A.N.).

(i) *Cloud No. 1.*—A group of three heads, situated over land about 20 miles north-west of Sydney, was observed at 1055 hours rising from the top of the general cloud layer. The centre head of the three was selected for detailed investigation. This head continued to grow (as shown in Fig. 1 (b)), reaching a maximum height of 11,100 ft. It then gradually subsided to 11,000 ft. in the next 3 minutes after which it completely subsided into the general layer in a further 4 minutes.
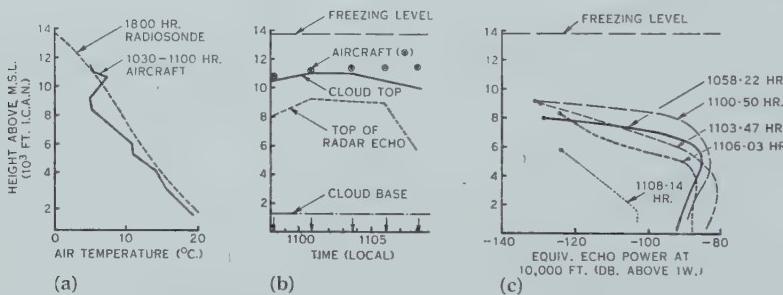
Fig. 1.—Observations of rain falling from non-freezing clouds on January 11, 1951 (Cloud No. 1).
    (a) Temperature sounding.
    (b) Heights of the aircraft, cloud top and base, and radar echo top.
    (c) Equivalent echo power.

During this period, five flights were made over the cloud at approximately $2\frac{1}{2}$-minute intervals, the rain falling from the cloud being observed on the aircraft radar set. A series of photographs of the most intense rain area observed during each of these runs is reproduced in Plate 2. The maximum echo intensity (Plate 2 (b)) was recorded during the second run which coincided with the time of maximum development of the cloud. At this time the height of the top of the radar echo coincided with a temperature inversion at 9200 ft. As soon as the cloud had passed maximum development and commenced to subside, the top of the echo began to fall (as shown in Fig. 1 (b)) with corresponding reduction in the intensity and lateral dimensions of the rain area.

The variation with time of the intensity of the echo from the most intense rain area in the cloud is shown in Figure 1 (c). From these curves of equivalent echo power v. height, it can be clearly seen that rain was still developing when first observed, reaching a maximum at the time of maximum development of the

cloud (1101 hours), thereafter gradually diminishing in intensity. The peak intensity during development of the rain occurred at an altitude of about 5000 ft., that is, in the vicinity of a shallow isothermal layer. During the dissipation stage of the cloud, the height of the peak echo as well as the height of the top of the echo diminished.

Maximum rain intensity on the ground occurred at 1103 hr. 47 sec., that is, 3 minutes after the cloud had started to subside. At this time, the rain intensity on the ground corresponding to the observed radar echo, was 70 mm./hr. which corresponds to " very excessive " rain as defined by Smith and Fletcher (1947). On the same basis, the average precipitation rate for the 10-minute period was about 35 mm./hr.

(ii) *Cloud No. 2.*—At 1125 hours the aircraft set course for an isolated formation of cumuli situated over the sea 5–10 miles east of Sydney. The head for investigation was rising from a circular area of cloud about 5 miles in diameter, the top of which was at 10,000 ft. This area of cloud was apparently the remnant of the 8/8 strato-cumuli which existed over the sea at 0900 hours.
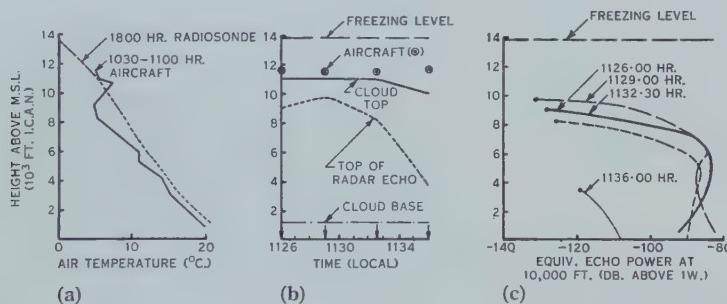

Fig. 2.—Observations of rain falling from non-freezing clouds on January 11, 1951 (Cloud No. 2).

    (*a*) Temperature sounding.
    (*b*) Heights of the aircraft, cloud top and base, and radar echo top.
    (*c*) Equivalent echo power.

During the first three runs over this head (as shown in Fig. 2 (*b*)) the top remained at 11,000 ft. On the last of these runs, however, it was dissipating and by the fourth run the top had fallen to 10,000 ft.

The maximum rain echo observed during each of these runs is shown in Plate 3. As can be seen, the rain echo grew in intensity until 1129 hours after which it rapidly diminished until, 7 minutes later, only a small echo remained. The variation in the height of the top of the rain echo with time is shown in Figure 2 (*b*), maximum height being 9700 ft. which was 500 ft. higher than in cloud No. 1 but was still limited by the temperature inversion. Although in this case the maximum development of the rain storm cannot be so definitely correlated with a peak in the development on the cloud top, it certainly did occur while the cloud was still active.

The curves of equivalent echo power *v*. height given in Figure 2 (*c*) again clearly show the growth and decay of the rain echo. The peak in the 1129-hours

echo power curve corresponds to the broad band of echo visible on Plate 3 (*b*) which lies between 6600 and 8300 ft. Three and a half minutes later the intensity was more uniform with height but had its maximum value between 5500 and 6500 ft. Neither of these maxima can be explained satisfactorily in terms of the temperature sounding. The maximum rain intensity on the surface of the sea was calculated to be 100 mm./hr. (1129 hours). The average precipitation rate over the period of 10 minutes was about 45 mm./hr.

(b) *January 23, 1951*

Maritime air was moving over the Sydney coast from a southerly direction in which at 1400 hours 8/8 strato-cumuli had formed over the land, with base at 4500 ft. and tops held down by inversion at 6000 ft. However, over the sea some isolated cumuli had developed with bases at 1750 ft. and tops varying from 10,000 to 12,500 ft. A temperature sounding made during the aircraft ascent is shown in Figure 3 (*a*) superimposed on the 1800-hours radiosonde sounding taken at Rathmines. Freezing level obtained from both soundings was 12,000 ft. (I.C.A.N.).
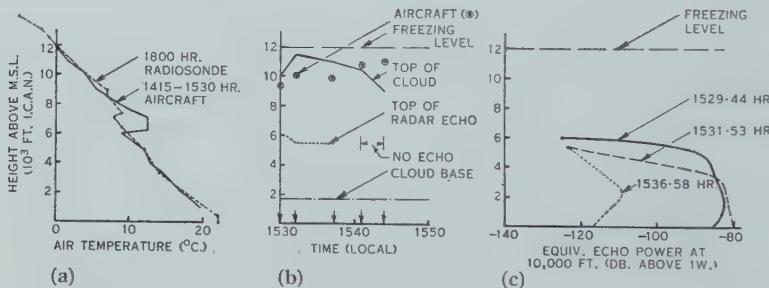

Fig. 3.—Observations of rain falling from non-freezing clouds on January 23, 1951.
   (*a*) Temperature sounding.
   (*b*) Heights of the aircraft, cloud top and base, and radar echo top.
   (*c*) Equivalent echo power.

A formation of cumuli was selected for special investigation whose top at 1530 hours was at 10,000 ft. and growing (Fig. 3 (*b*)). It reached its maximum development $2\frac{1}{4}$ minutes later. After a further 5 minutes the top had dropped from its greatest height to 11,000 ft. and was dissipating. Dissipation continued until the whole top had completely evaporated down to 9000 ft. The first three runs were made through the cloud; one at 9300, and two at 10,000 ft. On the first two runs, the cloud could be classed as " very wet " for a cumulus cloud, the size of the drops hitting the windscreen being estimated to be about $\frac{1}{4}$ mm. By the time of the third run, the water content had dropped to about half but the drop size was approximately the same. On none of these runs, as would be expected since the whole cloud was warmer than freezing, was any snow or ice observed. The turbulence encountered on the first two runs through the cloud (about 5 miles wide in an east-west direction) was moderate, maximum vertical accelerations experienced being of the order of $\pm\frac{1}{2}g$.

RAIN FROM NON-FREEZING CLOUDS
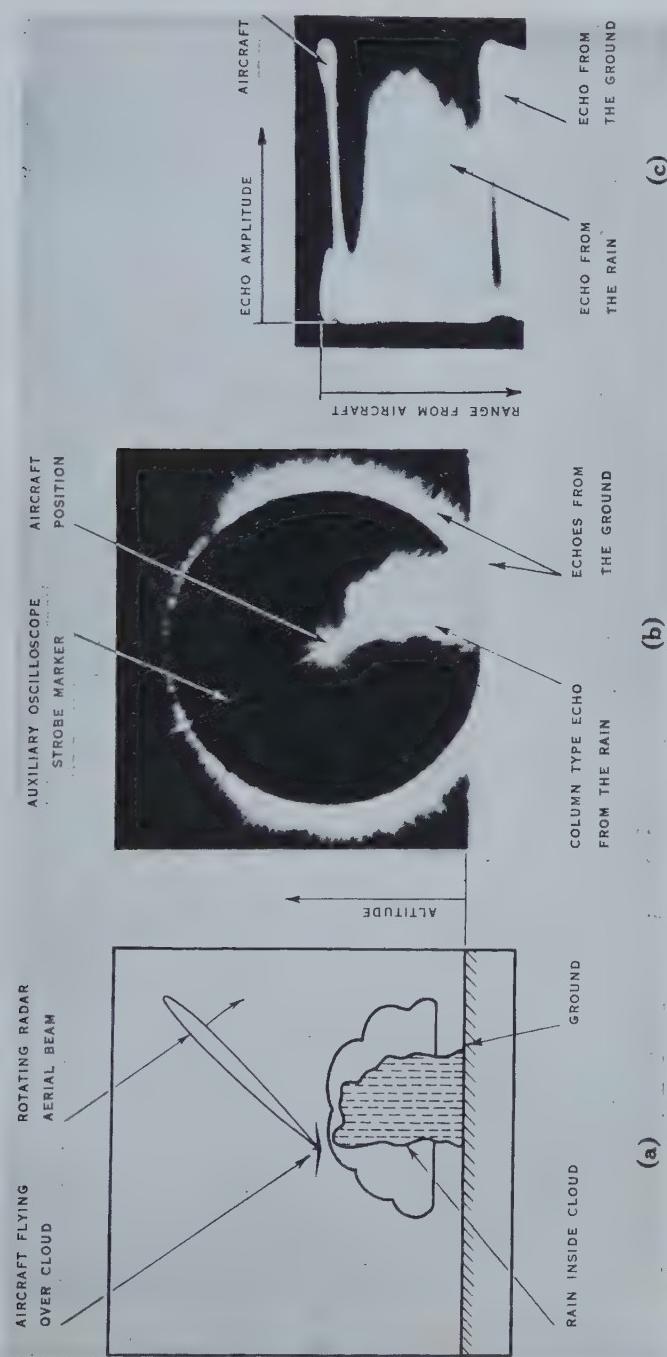
Diagram illustrating the principle of the airborne radar display.

(a) Aircraft fitted with the radar set is depicted flying above a raining cloud.
(b) The corresponding display of the radar echoes from the rain on a position indicator oscilloscope.
(c) The corresponding radar echo intensity down through the rain displayed on an amplitude-range oscilloscope.

RAIN FROM NON-FREEZING CLOUDS
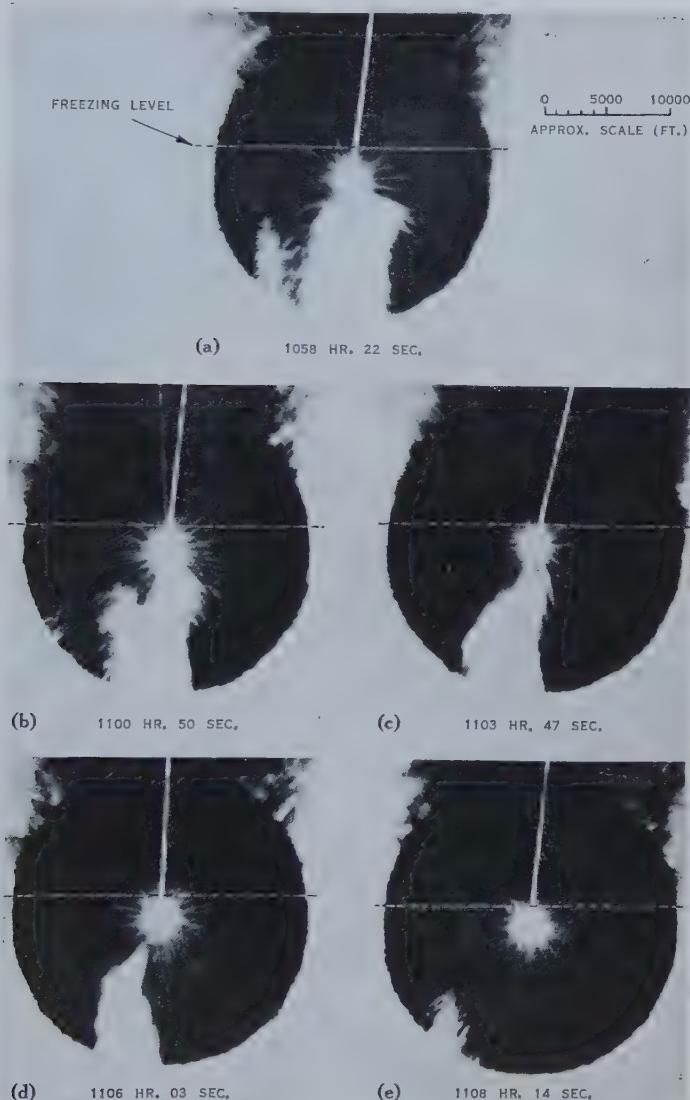
Airborne radar echoes from rain falling from non-freezing cloud on
January 11, 1951 (Cloud No. 1).

RAIN FROM NON-FREEZING CLOUDS
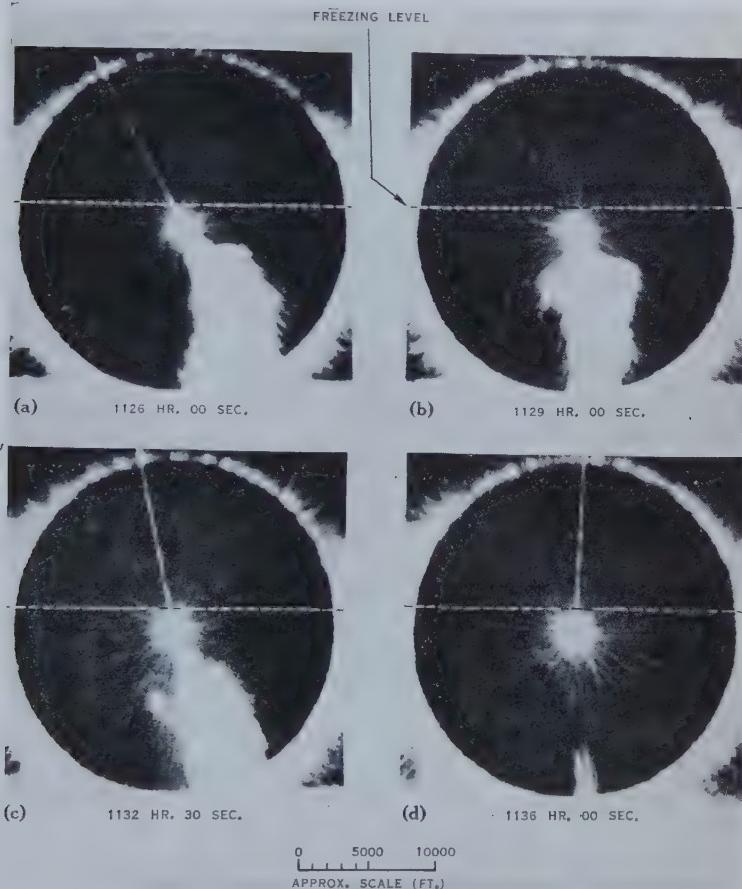
Airborne radar echoes from rain falling from non-freezing cloud on January 11, 1951 (Cloud No. 2).

RAIN FROM NON-FREEZING CLOUDS
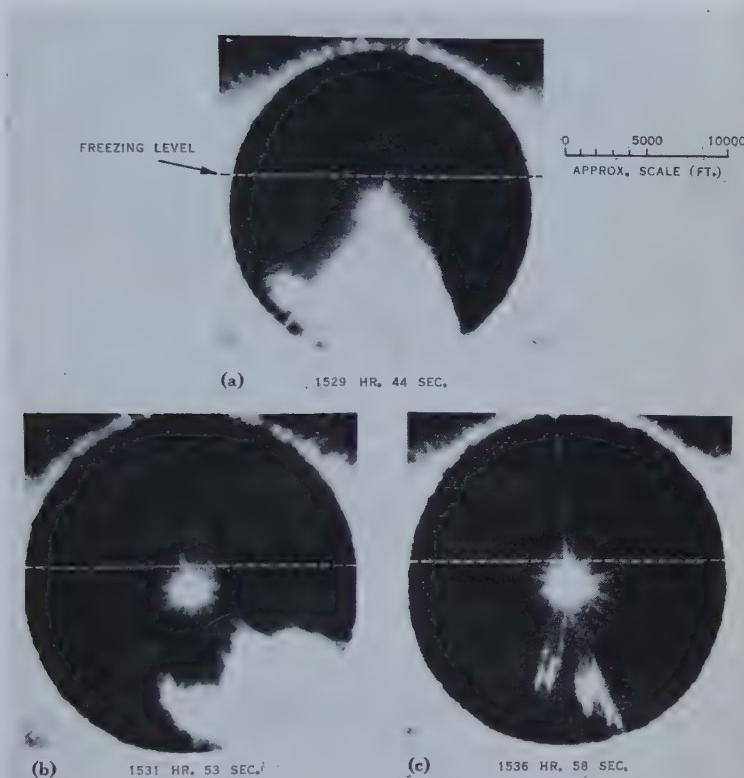
Airborne radar echoes from rain falling from non-freezing cloud on January 23, 1951.

RAIN FROM NON-FREEZING CLOUDS
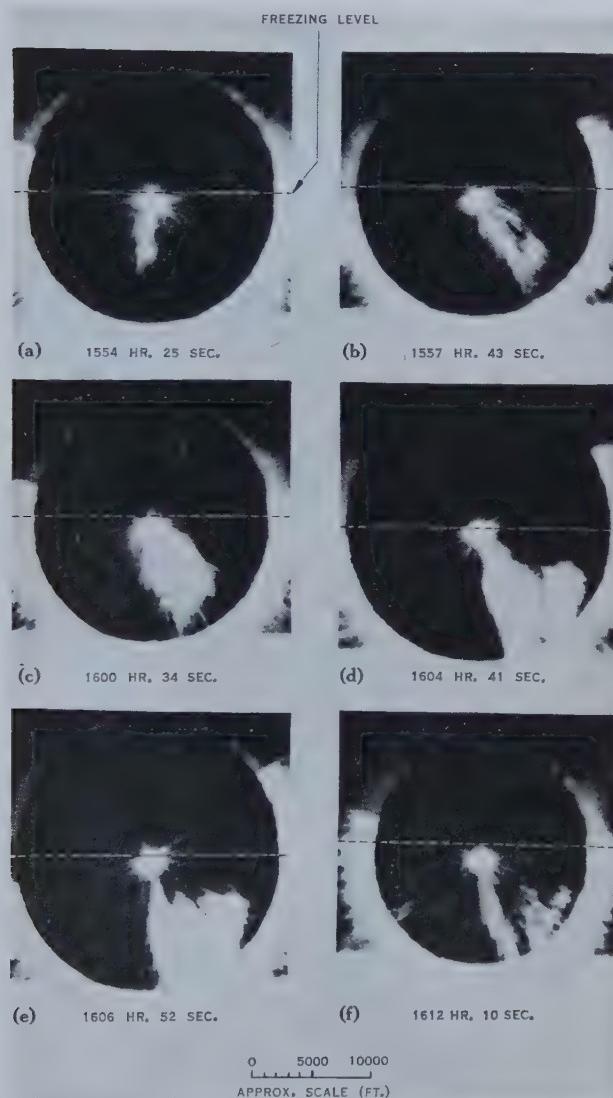
Airborne radar echoes from rain falling from non-freezing cloud on
January 24, 1951.

RAIN FROM NON-FREEZING CLOUDS
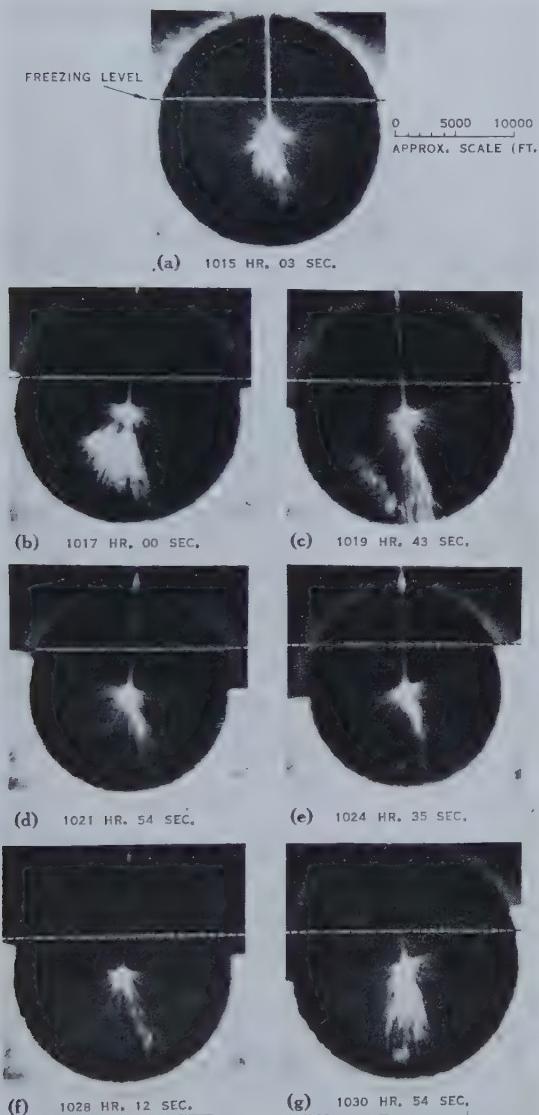
Airborne radar echoes from rain falling from non-freezing
cloud on March 22, 1951.

The photographs in Plate 4 show the most intense rain echo which was observed on each of these runs. The echo was strongest at 1529 hr. 44 sec. (Plate 4 (a)), that is, when the cloud was still actively building and not when the cloud had reached its peak development as was the case on January 11, 1951. The echo had nearly vanished 7 minutes later and completely so after 11 minutes.

The curves of equivalent echo power v. height shown in Figure 3 (c) confirm the fact that the most intense echo occurred at 1529 hr. 44 sec. after which it rapidly diminished. An interesting feature of this rain is the fact that the top of the rain echo at the time of its maximum development was held down at 6000 ft. by the $3\frac{1}{2}$ C. degree inversion and had fallen to 5500 ft. by the time the cloud had reached maximum development. Maximum rainfall intensity on the surface of the sea occurred at 1532 hours and was calculated to be about 100 mm./hr. The average precipitation rate was about 70 mm./hr. over the 7 minute period.

(c) January 24, 1951

A gentle maritime stream, associated with a weak cyclone centred about 200 miles east of Sydney, was moving across the coast from a southerly direction. Isolated cumuli were forming particularly over the sea, with bases at 1500 ft. and tops varying from 3000 to 20,000 ft.
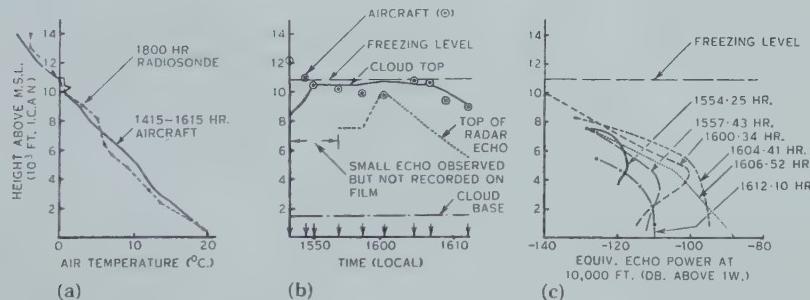

Fig. 4.—Observations of rain falling from non-freezing clouds on January 24, 1951.
    (a) Temperature sounding.
    (b) Heights of the aircraft, cloud top and base, and radar echo top.
    (c) Equivalent echo power.

In the afternoon a flight was made to investigate these clouds. A temperature sounding taken during the aircraft ascent is shown in Figure 4 (a) superimposed on the sounding obtained from the 1800-hours radiosonde release at Rathmines. Freezing level during the flight was 10,900 ft. whereas by 1800 hours it had dropped to 10,600 ft.

A line of cumuli was observed to be developing about 20 miles north-east of Sydney. The particular cloud selected for investigation was growing at 1548 hours, its top then estimated to be 8000–9000 ft. The development of this cloud in the period of observation is shown in Figure 4 (b). The cloud continued to grow until its top reached 10,500 ft. after which it remained stable, but still active, for 7 minutes. The top then rose to 10,750 ft. ($+\frac{1}{2}$ °C.) in the next 3 minutes. The top had almost completely evaporated down to 10,500 ft.

· F

$2\frac{1}{4}$ minutes later leaving a compact, but almost flat top at this height. This situation persisted for another 3 minutes, when this, also, started to dissipate and this stage continued until the cloud had completely evaporated down to 9100 ft., with the remainder very broken.

Three flights were made through the top of the cloud while it was still active. On the first two (1554 and 1601 hours respectively) a moderate water content was observed, the size of drops hitting the windscreen being estimated at less than $\frac{1}{4}$ mm. On the third run, however, 1000 ft. below the top of the cloud at the time of its maximum development, the water content was larger, drop size now being about $\frac{1}{4}$ mm. The maximum vertical acceleration experienced on any of these runs was $\pm\frac{1}{2}g$.

During the first three flights over the cloud (Fig. 4 (b), 1448–1551 hours) small echoes were observed momentarily on the radar but were not of sufficient lateral extent to be recorded on the radar film. Thereafter, however, the development and subsequent decay of the rain echo is clearly shown in the sequence of photographs in Plate 5. Curves of equivalent echo power v. height are given in Figure 4 (c). The rain echo when first recorded lay between 7500 and 3500 ft. This rain echo increased in size and intensity, until its top had reached 10,000 ft. (see Fig. 4 (b)), and its bottom had just reached sea-level, the peak intensity being at 4500 ft. Note that this echo coincided with the peak development of the cloud. The height of the top of the echo then immediately started to fall, but the intensity of the echo increased between 4500 ft. and the sea until at 1604 hr. 41 sec. it was almost uniform between these heights. Maximum intensity at sea-level was reached at 1607 hours. The precipitation rate corresponding to this observed echo intensity at sea-level is 30 mm./hr., which corresponds to " heavy " rain. After this the intensity at sea-level decreased, the average for a 12-minute period being 15 mm./hr.

(d) March 22, 1951

A gentle slightly modified maritime stream from a north-westerly direction was moving across Sydney ; 0–4/8 strato-cumuli were forming with bases at 2500 ft. and tops varying from 3000 to 9000 ft. but nowhere reaching freezing level, which was at 12,200 ft. A temperature sounding taken during the aircraft ascent is shown in Figure 5 (a) superimposed on radiosonde soundings taken at Rathmines at 0600 hours and 1800 hours.

The cumulus cloud selected for special investigation formed part of a line of cumuli running parallel to the coast and was situated on a bearing of 120 ° (Mag.) from Sydney, 25 miles out to sea. The behaviour of this cloud during the period of observation is shown by Figure 5 (b). It reached its maximum development (9000 ft.) at 1017 hours, the top thereafter gradually falling in the next $7\frac{1}{2}$ minutes, to 8200 ft. It then started to grow again, reaching 9200 ft. $6\frac{1}{2}$ minutes later, after which it subsided until the top was at 8500 ft. At 1022 hours and 1025 hours flights were made through the cloud about 100 ft. below its top. Only a low liquid water content was observed. Drops were only just visible on the windscreen, hence the drop size lay in a range up to $\frac{1}{4}$ mm. diameter. The maximum vertical accelerations experienced on these two runs through the

cloud were approximately $\pm\frac{1}{4}g$ and $\pm\frac{1}{2}g$ respectively. The larger acceleration experienced on the latter run can be accounted for by the fact that the cloud was starting to grow again. A flight made beneath the cloud base at 2500 ft. at 1055 hours revealed light rain, the drop size lying in the range $\frac{1}{2}$–1 mm.

A series of photographs of the radar echo from rain falling from the cloud are shown in Plate 6. This example of non-freezing rain is interesting, not because of the intensity of the rain, which in this case was light, but because of the effectively cyclic variation of the height of the top of the rain echo with the development of the cloud. This effect can readily be distinguished on the successive photographs in Plate 6, but is clearer when these are studied in conjunction with the corresponding equivalent echo power-height curves shown in Figure 5 (c). When first observed the echo was near the top of the cloud, height of top of echo and of the peak echo being at 8200 and 6250 ft. respectively. The
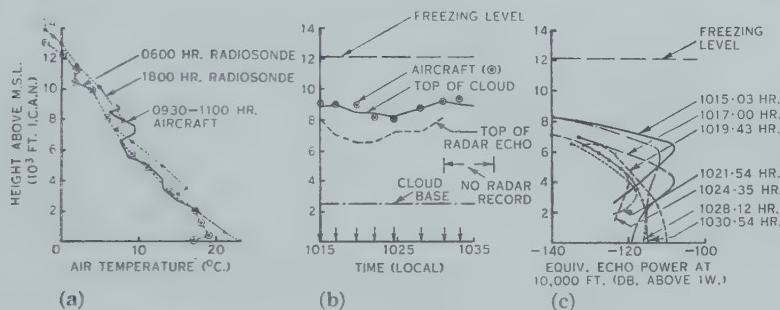
Fig. 5.—Observations of rain falling from non-freezing cloud on March 22, 1951.

(a) Temperature sounding.
(b) Heights of the aircraft, cloud top and base, and radar echo top.
(c) Equivalent echo power.

echo did not reach the sea. When the cloud was at maximum development for the first time (1017 hours) the top of the echo and peak echo had fallen to 7000 and 4000 ft. respectively, the rain still not reaching the sea. After 3 minutes, the top had fallen to 6500 ft. with maximum intensity close to the sea. The next cycle started 2 minutes later when the rain shower commenced to develop again, giving a small echo with maximum intensity at about 3000 ft., although the cloud top still appeared to be subsiding. After a further 3 minutes, when the cloud started to rebuild, the top of the echo had risen to 7200 ft. with its peak at about 4000 ft. The top of the echo then remained steady at 7200 ft. but the echo intensified between 3500 ft. and the sea. By 1031 hours, when the cloud had reached its maximum development a second time, the top of the rain echo had risen to 8200 ft. again with its peak intensity at 5500 ft. Unfortunately no further radar record was obtained owing to a fault in the camera.

However, it is confirmed again in this example that the maximum rain intensity occurs either when the cloud is still actively building, that is, prior to it attaining its maximum development, or when the cloud just reaches maturity.

The maximum echo intensity occurred on the surface of the sea at 1020 hours. The corresponding precipitation rate was 1 mm./hr., i.e. light rain, and averaged only 0·5 mm./hr. over the 5-minute period.

IV. SUMMARY OF OBSERVATIONS

The observations of natural rain described in this paper contain several important features (Table 1).

(*a*) The clouds were wholly warmer than 0 °C., the temperature of the top at the time of their maximum development being between $+7$ and $+\frac{1}{2}$ °C. in all cases.

(*b*) Calculations of precipitation rate on the ground based on the intensity of the radar echo and Laws and Parsons drop size distribution show that the rain intensity on the ground averaged over the period of the storm can reach 70 mm./hr. i.e. " excessive " proportions. Three cases of excessive rain are recorded, one of heavy, and one of drizzle.

(*c*) In the four cases of heaviest rain, the cloud formed in maritime air masses moving over the Sydney coast from a south or south-easterly direction. In the other case (March 22) the air mass was slightly modified maritime air moving back over the sea from a north-westerly direction.

In all cases the clouds investigated were either over the sea or only slightly inland.

(*d*) In the four examples where the precipitation rate was heaviest, the cloud depth was between 9250 and 9900 ft. The light rain of March 22 fell from a cloud only 6500 ft. deep. However, since the air mass was also different on this occasion no conclusion can be reached.

(*e*) The values of effective cloud life given in Table 1 are the times from when the head was first observed to commence its growth from the top of the general cloud layer to when it subsided back into the layer or had nearly completely dissipated. The clouds may hence be classed as " of short life ".

(*f*) The observed echo from the rain in the cloud was in the form of a vertical column.

(*g*) In all cases the maximum development of the radar echo occurred when the cloud was still actively building just prior to the cloud reaching maximum development or when the cloud was actually at maximum development.

(*h*) Also in all cases the cloud commenced either to dissipate or subside immediately upon reaching maximum development. Simultaneously the top of the radar echo also started to fall.

(*i*) In most cases the greatest rainfall intensity on the ground occurred after the cloud had passed maximum development.

(*j*) In four of these observations it will be noted that, while the cloud was still developing, a maximum in the radar echo intensity from the rain occurred in the middle levels in the cloud, indicating an accumulation of relatively large drops at these levels. In most cases there was little or no discontinuity in temperature sounding to account for such an effect. This effect was not

TABLE 1

SUMMARY OF OBSERVATIONS

Date	Cloud	Freezing Level (ft. I.C.A.N.)	Temp. of Cloud Top at Maximum Development (°C.)	Air Mass	Cloud Location	Depth of Cloud at Maximum Development (ft.)	Approx. Effective Life of Cloud (min.)	Time of Maximum Height of Cloud Top (hr. sec.)	Time of Maximum Height of Radar Echo (hr. sec.)	Time of Maximum Precipitation on Ground or Sea (hr. sec.)	Calculated Precipitation Rate Maximum (mm./hr.)	Classification	Average (mm./hr.)	Classification	Approx. Duration of Rain on Ground or Sea (min.)
Jan. 11, 1951	No. 1	13,800	+5½	Maritime	Over land	9900	14	1100.50	1100.50	1103.47	70	Excessive	35	Excessive	10
Jan. 11, 1951	No. 2	13,800	+5½	Maritime	Over sea	9800	>12	1126.00 to 1132.30	1129.00	1129.00	100	Very excessive	45	Excessive	10
Jan. 23, 1951		12,000	+1	Maritime	Over sea	9750	15	1531.53	1529.44	1531.53	100	Very excessive	70	Excessive	7
Jan. 24, 1951		10,900	+½	Maritime	Over sea	9250	25	1600.34	1600.34	1606.52	30	Heavy	15	Heavy	12
Mar. 22, 1951		12,200	+7	Slightly modified maritime	Over sea	6500	Two cycles each 15	1017.00 1030.54	1015.03 1030.54	1019.43	1	Light	0·5	Drizzle	5

observed, however, in the example of January 23, owing, perhaps, to the late start of observations on this cloud. In this cloud also, the top of the radar echo was limited by a $3\frac{1}{2}$ C. degree inversion near the middle level of the cloud.

Generally, after the cloud had passed maximum development, the region of high concentration of water drops disappeared leaving the echo intensity more uniform with height.

V. Calculations of Raindrop Density and Rainwater Content

From the curve of observed radar echo intensity v. height for a particular rainstorm, it is possible to compute the variation of raindrop density and rainwater content down through the cloud on various assumptions about the rate of growth of the falling drops.
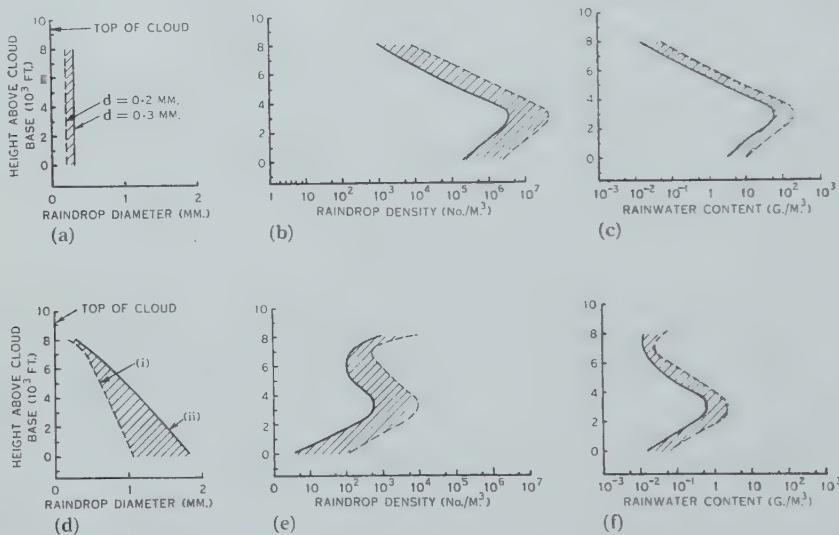
Fig. 6.—Curves of raindrop density and rainwater content derived from the observed echo intensity of the rain at 1600 hours January 24, 1951, assuming in (a)-(c) no growth of the falling drops and in (d)-(f) that growth occurs by coalescence. It is also assumed in the latter case that, for curve (i), the raindrop diameter d at 8000 ft. $= 0 \cdot 2$ mm., the average cloud droplet diameter $D = 20\mu$, the cloud water content $w = 0 \cdot 5$ g./m.³, and the upward air velocity $V = 70$ cm./sec. ; for curve (ii), d at 8000 ft. $= 0 \cdot 3$ mm., $D = 20\mu$, $w = 1$ g./m.³, and $V = 115$ cm./sec.

Let us consider the rainstorm for January 24, 1951, at the time of its maximum development. The data acquired upon which the following calculations were based, were, first, the equivalent radar echo power given in Figure 4 (c) (1600 hr. 34 sec.) and, second, an estimate of the drop size obtained during the flight through the cloud at a height of 8000 ft. above its base. The predominant drop size at this height was found to be $0 \cdot 25$ mm. in diameter, let us say in the range $0 \cdot 2 - 0 \cdot 3$ mm.

(i) If we now assume no growth of the raindrops as they fall through the cloud, then the diameter of the drops at every height in the cloud would be in this same range $0 \cdot 2 - 0 \cdot 3$ mm. (Fig. 6 (a)). The corresponding values of raindrop

density and rainwater content at different heights, plotted in Figures 6 (b) and (c), may then be calculated. It is immediately obvious that the values of both raindrop density (5×10^6 to 5×10^7 drops/m.³ at 3000 ft. above the cloud base) and rainwater content (70–200 g./m.³ at the same height) are absurd. For this reason, if for no other, the assumption that the raindrops do not grow in falling through the cloud must be invalid.

(ii) As an alternative, let us assume that in falling through the cloud the raindrops grow by coalescence at the rate given by the coalescence theory of Bowen (1950). On the further assumption that the upward air velocity, cloud water content, and average cloud drop diameter are as indicated in the legend to Figure 6, the drops would grow in the manner indicated in Figure 6 (d), attaining diameters between 1 and 2 mm. on reaching the cloud base.

The corresponding values of raindrop density and rainwater content at different heights are given in Figure 6 (e) and (f). The numerical values, although somewhat high, are not unreasonable.

It can therefore be stated with some degree of certainty that the raindrops must have grown in falling through the cloud and that the rate of growth was compatible with the coalescence theory.

VI. Conclusion

The rain described in this paper came from clouds that were wholly at temperatures above 0 °C. and the conditions were such that ice crystals could not possibly have been involved in the formation of the rain. It has already been shown (Bowen 1950) that such rain can form by coalescence and the observations described in the present paper are consistent with the theory.

The clouds were of moderate depth and formed in maritime air in the warm temperate region. The intensity of the rain at the ground varied from drizzle to one case in which the average intensity over the duration of the storm was 70 mm./hr.

VII. Acknowledgments

VIII. References

BOWEN, E. G. (1950).—*Aust. J. Sci. Res.* A **3** : 193.

HEYWOOD, G. S. P. (1940).—*Quart. J. R. Met. Soc.* **66** : 46.

HUNT, T. L. (1949).—*Met. Mag.* **78** : 26.

KOTSCH, W. J. (1947).—*Bull. Amer. Met. Soc.* **28** : 87.

LAWS, J. O., and PARSONS, D. A. (1943).—*Trans. Amer. Geophys. Un.* Part II, **23-24** : 452.

SIMPSON, G. C. (1941).—*Quart. J. R. Met. Soc.* **67** : 129.

SMITH, E. J. (1951).—*Quart. J. R. Met. Soc.* **77** : 33.

SMITH, E. O., and FLETCHER, R. G. (1947).—*Trans. Amer. Geophys. Un.* **28** : 713.

TWOMEY, S. (1952).—*J. Met.* (in press).

WARD, R. (1952).—*Met. Mag.* **81** : 56.

THE DETERMINATION OF SURFACE TENSION BY SESSILE DROP MEASUREMENTS, WITH APPLICATION TO MERCURY

By G. M. Ziesing*

Summary

The error in the Worthington (1885) sessile drop equation, as an approximation to the exact equation from which it was derived, is found to be of the order of experimental errors in recent surface tension determinations. The source of error and origin of the optical band system observed when parallel light is used to define the summit of a sessile drop have been investigated. The band system is shown to originate in an interference phenomenon from the curved surface of the drop. Detailed descriptions are given of the cleaning and testing methods necessary for the preparation of mercury and glassware involved in the determination of the surface tension of mercury as $484 \cdot 9 \pm 1 \cdot 8$ dyn./cm. A lower limit to the temperature coefficient of the surface tension was set at $0 \cdot 20$ dyn./cm. °C.

I. INTRODUCTION

The determination of the surface tension of a liquid from observations made on a large (sessile) drop resting on a horizontal surface has the advantage that readings are taken only on the liquid surface and do not involve angles of contact of liquid and solid. The method depends on a very precise measurement of h, the vertical distance between the summit of the drop and the horizontal plane of maximum section (Fig. 1), and a reading of the diameter, $2R$, which enters into the correction factor to be applied to the well-known formula for a drop of infinite diameter (i.e. $\gamma = \frac{1}{2}\rho g h^2$).

Workers applying this method to mercury and claiming a high degree of accuracy for their work have obtained values ranging from $436 \cdot 3$ to 515 dyn./cm. Apart from surface contamination of the mercury and the use of inadequate formulae these variations are probably due to errors in measuring h, caused firstly by the difficulty of setting on a highly reflecting surface such as the summit of the drop, and secondly by the fact that the microscope must be moved a distance R horizontally between setting on the " equator " and on the summit. Burdon (1932) discussed these points and values for the surface tension of mercury in more recent papers have been in much closer agreement. In his work, as well as the author's, the mercury drop rested on a slightly concave quartz disk enclosed in an apparatus made of clear silica (Fig. 3) having plane circular windows for illumination and observation. The plane of maximum section is defined precisely by the highly astigmatic reflection in the drop of a

* Department of Physics, University of Adelaide ; present address : Department of Supply, attached to Royal Aircraft Establishment, England.

distant small light on the same horizontal level, as used by Richards and Boyer (1921).

To define the summit Burdon set up a small light at the focus of a large lens of good quality at some distance on the far side of the apparatus so that the drop was silhouetted in the horizontal beam of parallel light. The reading microscope, resting on an accurately plane and horizontal slab of glass, was set on the equator and then pushed forward to focus on the summit. For the present work the apparatus has been modified to include a pointer fixed just above the position of the centre of the drop, so that the reading on the top of the drop is taken by setting midway between the point and its image in the surface,
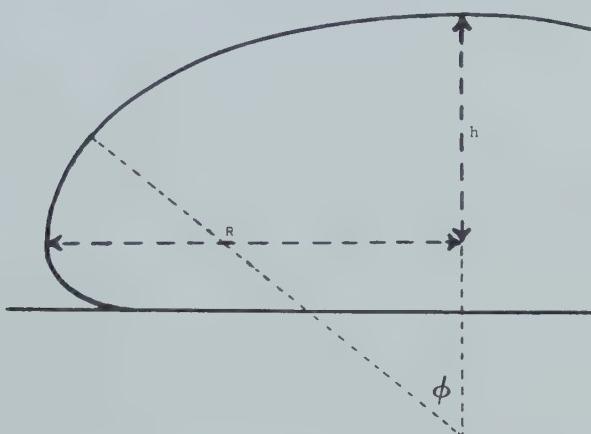
Fig. 1.—Section through sessile drop on horizontal plate.

as in the work of Bradley (1934) and Kemball (1946). Finding R to the required accuracy offers no difficulty. The present paper shows that modern technique warrants the use of a formula more accurate than the approximation used by previous workers in calculating the surface tension from their measurements. It also deals quantitatively with the optical interference fringes seen above the drop in the parallel beam of light and made use of by Burdon in setting on the summit. An account is given of a redetermination of the surface tension of mercury and its temperature coefficient.

II. FORMULAE APPLIED TO SESSILE DROPS

In recent determinations (Cook 1929; Kernaghan 1931; Burdon 1932; Bradley 1934; Kemball 1946) of the surface tension of mercury by measurements of sessile drops it has been customary to use the equation

$$\gamma = \tfrac{1}{2}[\rho g h^2 \times 1 \cdot 641 R/(1 \cdot 641 R + h)], \quad (R > 2 \text{ cm.}) \quad \dots \quad (1)$$

where γ = surface tension in dyn./cm.,
  ρ = density of mercury in g./cm.^{3},
  g = acceleration due to gravity,
  h = height of the drop measured above the plane of maximum section,
and R = radius of the drop at the plane of maximum section (Fig. 1).

The equation originally given by Worthington (1885) is

$$\gamma = \tfrac{1}{2}\rho gh^2 + 2\gamma h(1/b - 1/3 \cdot 282R), \quad \dotso \quad (2)$$

where b = radius of curvature at the summit of the drop.

The Worthington equation is itself an approximate form of the exact equation

$$\gamma = \tfrac{1}{2}\rho gh^2 + 2\gamma h/b - \gamma \int_0^h (\sin \varphi/x)\mathrm{d}x, \quad \dotso \quad (3)$$

where φ is the included angle between the axis of revolution and the normal to the surface (Fig. 1). Calculation shows that, for a drop of mercury $5 \cdot 0$ cm. in diameter, neglecting the value of $1/b$ causes an error of $0 \cdot 04$ per cent. and that equation (1) gives a value differing by $0 \cdot 23$ per cent. from that given by equation (3). The total effect is that values calculated using equation (1) are $0 \cdot 27$ per cent. (or $1 \cdot 2$ dyn./cm.) lower than those derived from (3). Nevertheless recent workers (Kemball 1946), whose measurements warrant the claim of $0 \cdot 3$ per cent. as their maximum error, have in fact used equation (1) thus involving a constant error approximately equal to the total estimated error.

It would appear that all results of workers using the sessile drop should be increased by at least 1 dyn./cm.

III. Optical Fringes observed above Sessile Drops

The method of using parallel light to define the summit introduces an error of $0 \cdot 001$ cm. due to the presence of what Burdon (1932) called " diffraction bands " above the liquid surface when viewed through the reading microscope. The nature of these diffraction bands and the magnitude of the resultant error in definition of the summit are discussed below.

Preliminary comparison of the parallel light method and the pointer reflection method of setting on the summit showed that the bands did not originate from the true summit of the drop, but from some vertical plane either on the near side or the far side of the summit. The plane of origin could be controlled by small variations of the angle of incidence of the parallel light on the drop, but, in spite of a fine control of movement, it was impossible to find a band system originating from the vertical plane through the summit of the drop.

Since a similar dependence on the angle of incidence was observed when the parallel light was replaced by light from a pinhole source, the phenomenon is attributable to the small departure from true parallelism which occurs with any optical system. The resultant error in h depends on the size of the drop on which the setting is made.

Apparently the origin of the unequally spaced optical fringes has not hitherto been described, apart from Burdon's reference to " diffraction fringes ". The fringes are characteristic of a liquid surface. The band spacing varies with variation of drop size, increasing as the drops become larger, which is contrary to measurements on diffraction fringes carried out by Arkadiew (1913)

and repeated in this laboratory.   The fringe spacing should be independent of change in radius of curvature of the object if the fringes are due to diffraction. A sharp edge, such as that of a razor blade placed normally to the beam in place of the mercury drop, produced no fringes if the microscope was correctly focused on the edge.

The knowledge that the bands or fringes being observed were actually formed in the same plane as the observed summit of the drop, a fact not explainable on normal diffraction theory, prompted further investigation.

A 3 in. optical flat, plane within λ/15, was set up horizontally with its upper surface in the position formerly occupied by the summit of the mercury drop, and with similar lighting it gave rise to an equally spaced system of fringes as in the arrangement known as Lloyd's single mirror (Fig. 2).   In such an arrangement the centre of the optical pattern is displaced from the plane of the optical
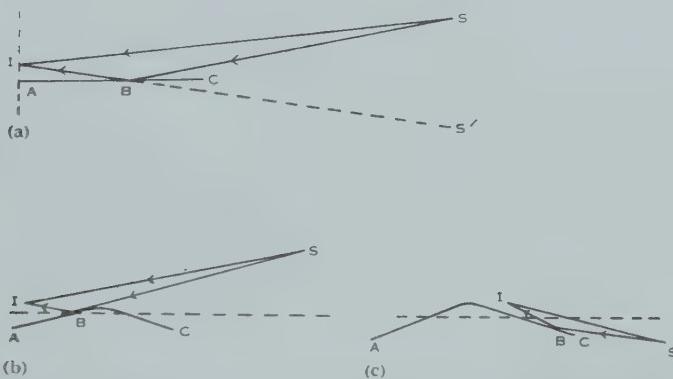
Fig. 2.—Interference occurs at I between the incident and reflected rays from S, either a distant pinhole, or with imperfectly parallel light.
  (a) From truly horizontal and flat surface (Lloyd's mirror) ;
  (b) From far side of curved surface ;
  (c) From near side.

flat due to the phase reversal of the reflected ray.   The plane of the optical flat is thus indefinite, as it has become the centre of a dark band.   If the plane of the flat is taken as the edge of the dark band, an error of half a bandwidth (0·0008 cm.) is involved.   The occurrence of the fringe system in the plane of the observed summit of the drop still required elucidation.

Since the upper surface of a sessile drop is not plane but has a large radius of curvature dependent on the drop size, the possibility of interference fringes from a curved reflecting surface was investigated.

In a paper entitled " The diffraction of light by cylinders of large radius ", where the maximum value of the radius was 1·5 cm., Basu (1918) considered the explanation of fringes formed when the focal plane of the observing instrument

coincided with the edge of the cylinder. A simple interference theory satis-
factorily accounted for his observed fringe spacing, which obeyed a law

$$x = \tfrac{3}{4}(2R)^{1/3}(n\lambda)^{2/3}, \quad n = 1, 2, 3 \ldots$$

where x=fringe spacing from minimum to minimum,
      R=radius of the cylinder,
and λ=wavelength of the light used.

If the liquid drop is considered as a cylinder of radius b, the radius of
curvature at the summit, reasonable agreement between predicted and measured
values of fringe spacings was obtained as shown in Table 1.

TABLE 1
RESULTS USING A WATER DROP OF RADIUS 3 CM.
Temperature=25 °C., $h = 0 \cdot 3996$ cm., $b = R = 940$ cm.*

Fringe number		1	2	3	4	5
Actual spacing..	..		0·0070	0·0058	0·0046	0·0042
Predicted spacing	..	0·0141	0·0082	0·0072	0·0062	0·0051

* Calculated by the formula of Laplace (1806) or Mathieu (1883).

The accuracy was limited as the fringe system did not originate exactly
at the summit, which would itself give a lower value for R (b) and hence better
agreement. Again, as R becomes very large the theory underlying the above
equation becomes inapplicable as the finite size of the source limits the definition
of a point as the edge of the cylinder. The formula predicted the increase of
band spacing with increase of drop size, since the radius of curvature increases
as drop size increases.

The fact that the fringes originated in an interference phenomenon and not
in diffraction was further confirmed by the effect of small vibrations on the
pattern. Minute ripples on the surface, insufficient to cause noticeable movement
of the liquid plane, caused the outer fringes to pulsate quite noticeably, owing
to the rhythmic variation of angle of the reflected light as the disturbance
crossed the surface.

The two errors in the parallel light method were in opposition, the fringe
error giving a constant high value to h, while the error due to the plane of origin
of the fringes varied with drop size. In the present experiments, for a water
drop of radius 3 cm., the effects cancelled out and no error existed between the
parallel light method and the pointer reflection method. For smaller radii the
error involved was negative, while at larger radii the error was positive, the
exact amount being uncertain to a degree determined by the accuracy of the
theory applied above. The direction and magnitude of the error were experi-
mentally confirmed. Thus the parallel light method is unreliable for accuracy
beyond $0 \cdot 001$ cm. in measurement of h.

When using the pointer reflection method the error in definition of h is
limited to observational error. Several papers (Taylor and Alexander 1944 ;
Wheeler, Tartar, and Lingafelter 1945) have been published in which the
technique applied in the pendant drop camera has been used for sessile drops
within the range of Bashforth and Adams's (1883) five figure tables of calculation.

If sufficient refinement can be introduced to either or both of these experimental arrangements, the order of accuracy will be beyond that of existing measurements and detailed knowledge of the surface tension variation of mercury at higher temperatures than previously measured should become possible.

IV. The Surface Tension of Mercury

Following preliminary experiments with a pendant drop camera in which it was not possible to obtain consistently repeatable results owing to the lens in use, the determination of the surface tension of mercury was made by means of large sessile drops in the silica apparatus formerly used by Burdon (Fig. 3). This apparatus was modified by the inclusion of a fixed pointer mounted within $0 \cdot 001$ cm. of the estimated position of the summit of a drop of mercury $5 \cdot 0$ cm.

TO McLEOD GAUGE,
CUT–OFF TRAPS,
AND PUMPS
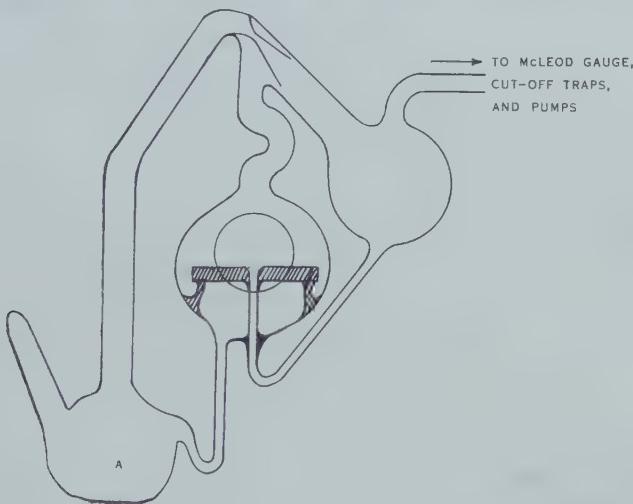
Fig. 3.—The silica apparatus.

in diameter at 25 °C. The vacuum apparatus consisted of a backing pump, mercury diffusion pump, liquid air trap, Pirani gauge, and mercury trap to isolate the greased side of the system from the other side, which comprised a McLeod gauge, charcoal trap, and silica measurement apparatus. A graded Pyrex-to-silica seal was made between the McLeod gauge and the charcoal trap. Provision for the inlet of clean dry gases was made near the diffusion pump such that all gas first passed through the liquid air trap. Each section of the vacuum apparatus was cleaned in chromic acid, washed in conductivity water (conductivity $< 3 \times 10^{-6}$ mhos), and dried before sealing together. No grease was applied to taps or cone joints until just before final sealing on of the main silica apparatus.

Special attention was paid to the cleaning of all glassware used in either the main or subsidiary operations. The general cleaning of all glassware used in subsidiary operations and not in contact with purified mercury followed the usual procedure of using tongs, washing in hot chromic acid, and rinsing in clean water. Two simple visual tests for cleanliness were applied to all glassware.

For the first, an unbroken water film indicated freedom from gross contamination. In the second test the article was tilted and the absence of an ascending water ridge known as the " Burdon " or " plate " ridge indicated the absence of partial contamination on the water surface covering the article (Burdon 1926 ; Satterly and Mills 1927 ; Satterly and Turnbull 1929 ; Satterly and Collingwood 1931a, 1931b ; Satterly and Stuckey 1932 ; Satterly and Givens 1933). When apparatus was required to be especially clean, the well-known technique of steaming was used. However, pilot tests had to be maintained on this process to safeguard against the possibility of airborne fatty contamination becoming interdispersed with the steam and coating the apparatus being steamed.

Where rapid drying was required, glassware was suspended on clean glass supports in a small silica oven, a process which was shown to be satisfactory.

The main silica apparatus was washed for 3 hr. in freshly prepared hot chromic acid, with acid being continually rinsed through it. The excess acid was washed off by 25 complete changes of conductivity water (conductivity $<2 \cdot 5 \times 10^{-6}$ mhos). The apparatus was steamed for 3 hr., dried in the silica furnace, and attached to the vacuum apparatus, which was then pumped and flamed for 9 hr. to remove as much adsorbed water vapour as possible.

The cleaning of the experimental mercury started with the preliminary experiments and continued until the vacuum apparatus had been completed. The starting sample was 400 cm.^{3} of stock mercury which had been freshly distilled at 200 °C. in a current of air. This was then passed through a series of two nitric acid/mercurous nitrate towers many times, being collected each time in cleaned and steamed glassware. The sample was then agitated in shallow Pyrex dishes under conductivity water to which pure sodium peroxide was continually added in small quantities. This final process occupied 60 hr. The purity of the final sample was tested by the spreading of conductivity water on a newly formed surface as outlined by Burdon (1926).

Since there is no adsorption of mercury by water or vice versa, Antonoff's law for the case of pure water on mercury reduced to the statement that no spreading would occur provided no ions were present in the water to become attached to the mercury surface and act as anchorages facilitating spreading. With water of known purity, the test was very sensitive as regards mercury purity. Conductivity water (conductivity $<2 \times 10^{-6}$ mhos) did not spread on the initial air distilled sample, nor did it spread on the sample obtained from the acid towers. After 18 hr. under sodium peroxide, limited spreading occurred, becoming a maximum after 60 hr. when water with conductivity $<1 \cdot 4 \times 10^{-6}$ mhos spread to cover a circle of diameter $5 \cdot 9$ cm. in 3 min. This drop eventually covered the whole area of the petri dish (except for the curved edge (Burdon, 1926)) before evaporating. Exposure of the mercury surface, protected from dust, to the air for half an hour did not alter the rate of spreading. These tests indicated a high standard of purity.

During these tests several observations previously made by Burdon were confirmed. Mercury was very liable to pick up contamination (possibly sodium ions) from dishes which had been thoroughly cleaned and steamed. Quite often conductivity water would not spread on the first mercury surface formed in a

clean dry dish, but the second surface formed in the same dish after rinsing with mercury was perfect, with spreading equal to the best tests performed. As Burdon had shown, adsorbed sodium ions prevented the spread of water on mercury, so that all mercury was stored in silica flasks which did not introduce this contamination.

With this cleaning complete the vacuum apparatus was filled with dry argon, opened, and the clean dry mercury poured into the reservoir chamber (A, Fig.3) from a special flask, constructed so that all mercury delivered came from beneath the free surface in the flask, ensuring absence of contamination. When the apparatus was resealed the argon reduced the risk of mercury oxidation due to the high temperature flame used.

At this stage the error introduced by the observation window not having truly parallel faces, mentioned by Kemball (1946), was finally checked. The wedge angle was $1 \cdot 0 \times 10^{-3}$ rad., which for the apparatus in use meant that the measured value of h was too high by $0 \cdot 0011$ cm.

TABLE 2

EXPERIMENTAL ERRORS

Type of Error	Error in h (cm.)	Error in γ (dyn./cm.)
Height of the lamp defining the equator..	$0 \cdot 0001$	$0 \cdot 3$
Artificial horizontal plane	$0 \cdot 0001(2)$	$0 \cdot 4$
Diameter of the drop	—	$0 \cdot 1$
Sum of errors		$0 \cdot 8$

Mercury was distilled around the apparatus for several days at gas pressures of less than 10^{-5} mm., which, according to the values of Taylor and Hulett (1913). were sufficiently far removed from the critical pressure for oxide formation to avoid the production of an oxide layer. When the final formation of a sessile drop had been made by distilling mercury from the reservoir chamber, with a small bunsen flame, a thermostat chamber maintaining a constant temperature of $25 \cdot 0 \pm 0 \cdot 1$ °C. was placed around the apparatus. Measurement of the surface tension was made when the temperature as registered by a chromel-alumel thermocouple and potentiometer had been constant for 2 hr.

The distillation of mercury within the silica apparatus gave rise to large electrical charges on the sessile drops, which depressed the value of γ by as much as 6 dyn./cm. A small dental X-ray plant was used to restore the original, earth, potential, thus ensuring the normal value of the surface tension. This was always carried out as soon as the drop was at 25 °C.

The value of the surface tension remained constant for 24 hr. The mean of the 42 readings taken was $483 \cdot 6 \pm 1$ dyn./cm., when calculated by the Worthington (1885) formula. The experimental errors were as set out in Table 2 so that the final result was $483 \cdot 6 \pm 1 \cdot 8$ dyn./cm. on the Worthington formula, which when corrected for the inaccuracies in that equation became $484 \cdot 9 \pm 1 \cdot 8$ dyn./cm.

V. The Temperature Coefficient of the Surface Tension of Mercury

Two series of values of surface tension against temperature were measured using a specially constructed furnace surrounding the silica apparatus. Within the furnace the formation of sessile drops continued by slow distillation, so that the results represent the surface tension of a series of slowly formed new surfaces. However, the electrical charge on the drop was also being continually built up as the distillation proceeded, so that the measurements made were limited in value as exact measurements. They did have value in one sense, for, since the influence of charge was greater at low temperatures where electrical equilibrium was harder to maintain, the values set a lower limit to the temperature coefficient of the surface tension of mercury. The measured value was $0 \cdot 20 \pm 0 \cdot 01$ dyn./cm. °C. over the range 20–220 °C. The uncertainty in this value demonstrates the limitation in measurement of a temperature coefficient by means of sessile drops, since, even under the best conditions, at any temperature the observational error is generally ± 1 dyn./cm., so that, unless the temperature range can be extended or an improved order of accuracy introduced, the value of the temperature coefficient has a lower accuracy of only $0 \cdot 01$ dyn./cm. °C. The use of a camera is the most promising method in both cases.

VI. Conclusion

The experimental value for the surface tension of mercury was in good agreement with previous work as reviewed by Kemball (1946), especially when all sessile drop measurements in that review were corrected for the inaccuracy in the Worthington equation. The value for the temperature coefficient was a good lower limit to the values quoted by Kemball.

VII. Acknowledgment

This paper describes part of the work for an M.Sc. thesis and was carried out, under the direction of Dr. R. S. Burdon, in the Physical Laboratory, Department of Physics, University of Adelaide.

VIII. References

ARKADIEW, W. (1913).—*Phys. Z.* **14** : 832.

BASHFORTH, F., and ADAMS, J. C. (1883).—" An Attempt to Test the Theories of Capillary Action." (Cambridge Univ. Press.)

BASU, N. (1918).—*Phil. Mag.* **35** : 79.

BRADLEY, R. S. (1934).—*J. Phys. Chem.* **38** : 231.

BURDON, R. S. (1926).—*Proc. Phys. Soc. Lond.* **38** : 148.

BURDON, R. S. (1932).—*Trans. Faraday Soc.* **28** : 866.

COOK, S. G. (1929).—*Phys. Rev.* **34** : 513.

KEMBALL, C. (1946).—*Trans. Faraday Soc.* **42** : 526.

KERNAGHAN, MARIE (1931).—*Phys. Rev.* **37** : 990.

LAPLACE, P. S. (1806).—" Mécanique Céleste." 2me Supplém. au livre 10, p. 483.

MATHIEU, E. (1883).—" Théorie de la Capillarité." p. 137. (Gauthier-Villars : Paris.)

RICHARDS, T. W., and BOYER, S. (1921).—*J. Amer. Chem. Soc.* **43** : 274.

SATTERLY, J., and COLLINGWOOD, L. H. (1931a).—*Trans. Roy. Soc. Can.* **25** (3) : 205.

SATTERLY, J., and COLLINGWOOD, L. H. (1931b).—*Trans. Roy. Soc. Can.* **25** (3) : 215.
SATTERLY, J., and GIVENS, H. (1933).—*Trans. Roy. Soc. Can.* **27** (3) : 145.
SATTERLY, J., and MILLS, W. J. P. (1927).—*Trans. Roy. Soc. Can.* **21** (3) : 277.
SATTERLY, J., and STUCKEY, E. L. (1932).—*Trans. Roy. Soc. Can.* **26** (3) : 131.
SATTERLY, J., and TURNBULL, R. (1929).—*Trans. Roy. Soc. Can.* **23** (3) : 95.
TAYLOR, G. B., and HULETT, G. A. (1913).—*J. Phys. Chem.* **17** : 565.
TAYLOR, H. J., and ALEXANDER, J. (1944).—*Proc. Indian Acad. Sci.* A **19** : 149.
WHEELER, O. L., TARTAR, H. V., and LINGAFELTER, E. C. (1945).—*J. Amer. Chem. Soc.* **67** : 2115.
WORTHINGTON, A. M. (1885).—*Phil. Mag.* **20** : 51.

A DECIMAL COUNTER ELECTRON TUBE

By D. L. Hollway*

Summary

A scale-of-10 counter electron tube is described, in which incoming signals are counted by triggering a single electron beam through a closed sequence of 10 pairs of stable states. Each stable state results from potential drops maintained on deflecting electrodes by the distribution of the beam current.

Current distributions and deflecting electrode systems of fivefold symmetry, which together develop rotating deflexion fields suitable for counters, are considered and an expression is derived for the optimum proportions of the electrode system.

Other possible applications of the basic principle, the " self-rotating beam ", to special purpose tubes are indicated.

I. Introduction

The need for counters operating at speeds higher than those possible with mechanical registers was realized first by nuclear physicists in recording signals from ionization chambers. An important advance was made when Wynn-Williams (1931) described a circuit of " a group of inter-connected thyratrons, which automatically switch one another into operation as the need arises, so that however rapidly particles are arriving, there is always a thyratron available for each particle ".

In later forms, electronic counters are now widely used in the measurement of frequency, time, and other quantities which may be converted to a frequency or count.

The chief requirement of any counter, whether mechanical or electronic, is a form of " memory " consisting of stable states, positions, or potential levels. It may be shown that at least $2N$ stable states are needed for a counter in the scale-of-N ; of these, N store the count in the intervals between incoming signals and the remainder hold the count during changeover. In mechanical counters such as cyclometers, clock escapements, and so on, it is simple to design the mechanism so that all of the $2N$ states are stable during any period of time. But the equivalent " completely stable " states in electronic counters require vacuum tubes or other comparatively costly " active " circuit elements. It is usual, therefore, to simplify these circuits by using " temporarily stable " states for storage during changeover. The crossed grid-to-plate condensers in the four double-triode scale-of-10 circuits are an example of this economy. Input shaping circuits ensure that the incoming signals are completed before the

* Division of Electrotechnology, C.S.I.R.O., University Grounds, Sydney.

stored charges decay.   These counters represent such an efficient use of com-
ponents that little further simplification seems possible in designs based on the
use of conventional vacuum tubes.

In 1947, in collaboration with Mr. A. M. Thompson, the possibility of
devising a single high-vacuum-tube decimal counter was considered by the
author, and several forms were planned.*   The objective of the subsequent
development was a counter of small size, operating at low anode voltage, with a
minimum of associated circuitry, which would maintain counting accuracy over
a wide range of circuit conditions and to as high a frequency as possible.

II. Principle of Operation
(a) General Description

The counter is a high-vacuum tube in which counting is accomplished by
successive displacements of a deflected electron beam through a number of
stable positions (Hollway 1950).   The operating principle may be explained by
considering the simplified electrode arrangement shown in Figure 1.
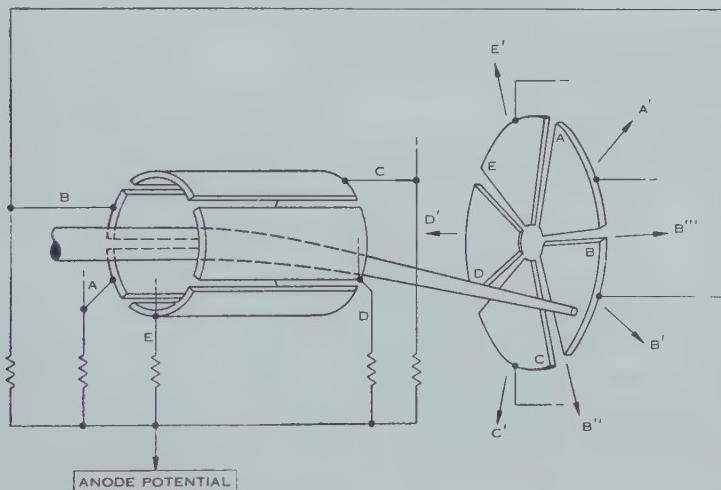
Fig. 1.—A simplified electrode system.

An electron beam enters, along the axis, a system of five similar deflector
plates maintained near the potential of the final anode by five equal series
resistances.   The beam forms a focused spot in the plane of the collecting
segments, each one of which is directly connected to the correspondingly lettered
deflector.

If one deflector, B, is lowered momentarily in potential, an electrostatic
field is formed which deflects the beam away from deflector plate B and moves
the spot in the direction of the arrow B'.   The beam current which then flows
to collector B will lower the potential of the collector, and hence of the deflector,

* It was found later that one of these early designs had been anticipated by a form of counter
proposed by Snyder (1946).

sufficiently to maintain the deflexion. A similar stable condition is possible at each of the other collectors, after initial deflexion.

Now imagine the deflector plates to be turned about the axis one-tenth of one revolution clockwise, viewed in the direction of the beam. Deflexion of the spot in the new direction, B'', will cause part of the beam current to flow to collector C, which in turn moves the spot further across plate C. There are no stable regions and the spot rotates continuously in the clockwise direction. Similarly, if the deflectors are retarded in space phase with respect to the collectors, to B''', the motion is reversed.
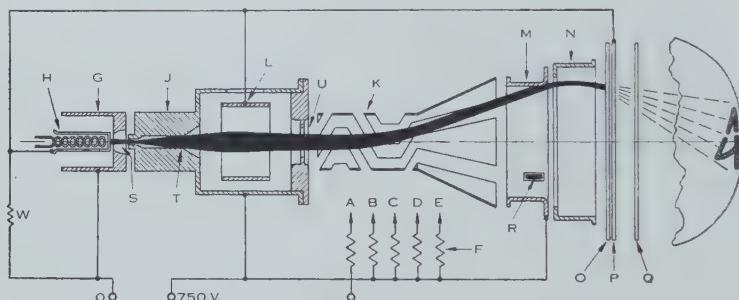
Fig. 2.—The electrode structure of the scale-of-10 counter, shown in section.
F, resistors connected to the collector plates A-E (shown in Figs. 4 and 5) ; G, grid ; H, cathode ; J, anode ; K, five deflectors ; L, focusing electrode ; M, positive ring ; N, trigger ring ; O, front collector plates ; P, suppressor ; Q, back collector plates ; R, carry-over electrode.

In the complete counter tube (Figs. 2 and 6) two separate sets of collector plates are used. The front group O is phased, with respect to the deflectors, for clockwise rotation and the back group Q is phased for reverse rotation. Openings are cut in the front group as shown in Figure 3. Thus, whenever the beam falls on the front plates it moves clockwise until part of the spot overlaps the radial edge of an opening and a fraction of the beam current flows to the back collectors shown in Figure 5. Further movement rapidly reduces the influence of the front system and increases that of the back, so that a state of angular equilibrium is reached, and the resultant deflexion is purely radial. Therefore the beam moves outwards until it is partially intercepted by the inside edge of a ring (M, Fig. 2), the " positive ring ", which is maintained at the potential of the final anode. Any further radial deflexion decreases the proportion of the beam which reaches the collectors, and the deflecting potentials are similarly reduced. Therefore a condition of equilibrium is reached when the beam is partly intercepted by the positive ring. Because the beam cross section is small in the plane of the ring, the radial deflexion remains almost unchanged by circuit variations or by the normal cycle of deflector potential changes which take place during the counting process.

In this way the angular position of the spot is always controlled by the deflector potentials and is sensitive to the collector currents, but these can have no effect on the radial position. The radial position is controlled only by the
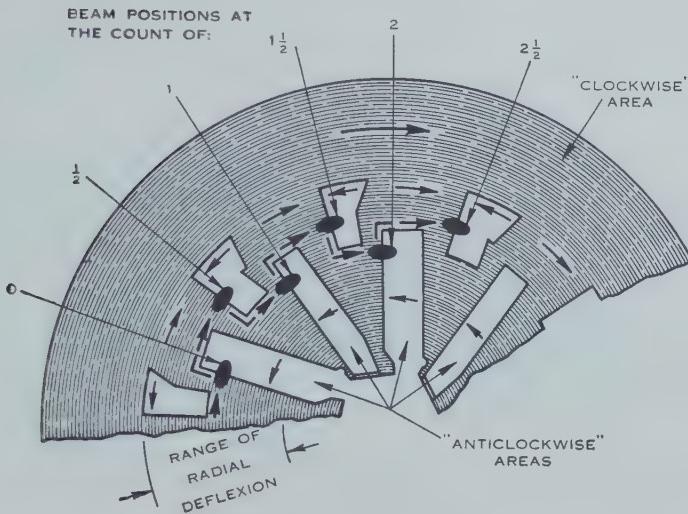
Fig. 3.—A simplified view of the stable openings in the front collectors showing the counting sequence 0→1→2.   Beam current falling on any part of the shaded area sets up potential changes which tend to move the spot clockwise.   This clockwise area consists of the five separate electrodes shown in Figure 4.
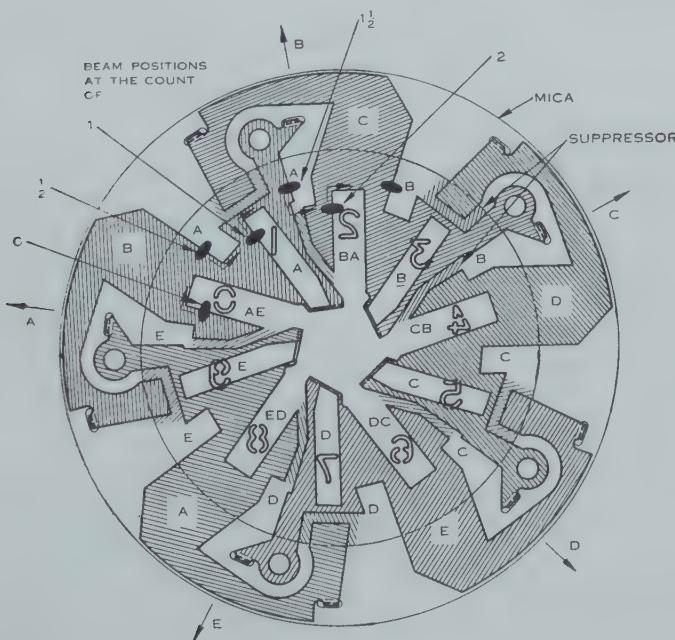
Fig. 4.—The collector assembly viewed in the direction of the beam, shown in relation to the directions of deflexion.   When beam current reaches a shaded area, for example B at position 0, the corresponding deflector is lowered in potential producing a component deflexion in the direction of the arrow B.

trigger electrode (*N*, Fig. 2), a short cylinder which receives the signals to be counted.

When a positive going signal appears, the spot moves radially outwards from the position 0 (Figs. 3 and 4) until the outer edge of the slot is reached. Beyond this point the anticlockwise restraint from the back collectors is reduced and the spot moves clockwise to the outer opening $\frac{1}{2}$, which is stable for this and higher trigger potentials. The beam is held in position $\frac{1}{2}$ until the removal of the input signal moves the spot inwards, once more on to a clockwise region, causing a further rotation to position 1.
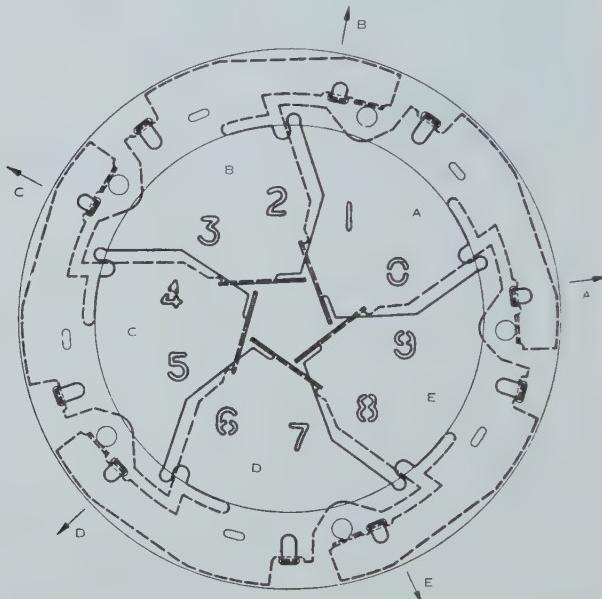
Fig. 5.—The back collector segments viewed from the screen against the direction of the electron beam.

Therefore each input signal moves the beam from one inner stable position to the next, increasing the stored count by one unit, and 10 waves cause the spot to complete a full revolution and return to the initial position. Because the states in which the beam is held during changeover are " completely stable ", the input signals may be of any duration and amplitude above the minimum values necessary to register a count.

Some interchange of secondary electrons is permitted between adjacent plates in both front and back groups in order to develop current waveforms corresponding to smoothly rotating deflexion fields, but complete suppression of secondaries between the two groups of collectors, and between the collectors and the other electrodes, is essential. This is ensured by keeping the trigger at a lower potential than the collectors, and by inserting a suppressor (*P*, Fig. 2) at cathode potential, between the front and back collectors. The suppressor has slots to match those in the front collectors. The outer 10 are larger than those in the collectors, so that the transmitted beam remains in good focus.

At the inner openings, however, the suppressor is formed so as to diverge the beam sufficiently to cover areas of the back collectors in which the figures 0–9 are cut, corresponding to the count positions. Electrons passing through the number openings continue to diverge, and project the number image,* enlarged eight times, centrally on the fluorescent end of the bulb.

Fig. 6.—A scale-of-10 counter tube.

When the beam reaches the position $9\frac{1}{2}$, the " carry-over " electrode (R, Fig. 2) intercepts the current usually flowing to the positive ring. This produces a negative pulse at R which is coupled to the trigger of the next counter as described in Section IV.

The principles underlying the division of the beam current between the collectors at each of the stable states will now be considered.

(b) The Deflector Potential Waveforms

In describing the operation of the counter it has been stated that the beam current falling at any point on the front group of collectors sets up a deflexion field which tends to move the spot further in the clockwise direction, that is, at any spot position, the current to the front group corresponds to a radial deflexion component at an angle, α, in advance of the actual position of the spot.

The values of α and the corresponding angle of the backward component depend upon the design of the collector plates and other electrode dimensions including the angular relationship between the collectors and the deflectors. Thus, variations within the electrode constructional tolerances cause variations in α, which the design value must be large enough to absorb in addition to the margin needed for protection against circuit and external disturbances. The tube will count incorrectly if either angle is reduced to zero in any region of the collector circumference.

As perturbations may occur with equal probability in the backward or forward directions and at any spot angle, it is assumed, as a design basis, that

---

* This method of number projection is preferable to an earlier form using electrostatic projection lenses, because it results in a simpler collector system and only the small fraction of the beam passing through the numbers is lost from the back plates ; also the images are always in sharp focus and satisfactory projection is maintained for a wider range of trigger potentials.

both components should be equal, equally spaced at $+$ and $-\alpha$ with respect to the direction of the spot, and that α should remain constant. By this concept the components follow the spot in its movement around the circumference, and thus the stability of all states is essentially the same.

In early experiments it was found that, in positions almost midway between the deflexion directions of two deflector plates, the angular position of the spot was sufficiently sensitive to differential changes in the potentials of only those two deflectors, to provide highly stable states. For this condition $\alpha \approx \pm 36°$. Too small a value of α means reduced stability, but if α is made unnecessarily large beam current is " wasted ", as a larger current is needed for a given radial deflexion.
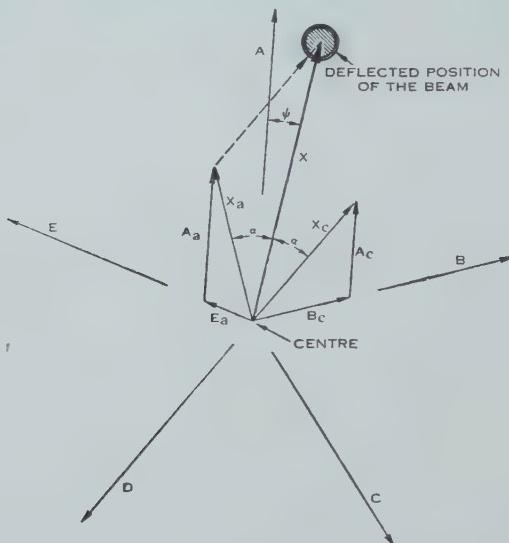
Fig. 7.—Components of the beam deflexion.

The individual collector currents needed to generate rotating fields of this kind are found from the vector diagram shown in Figure 7. The directions A to E are those of the five deflector plates ; that is, the spot would be moved in the direction A by making the potential of deflector A less positive The angle of the spot deflexion vector $\bar{X}$, is ψ, measured to A as a reference. The clockwise and anticlockwise components of $\bar{X}$, $\bar{X}_c$ and $\bar{X}_a$, therefore fall at angles $\psi + \alpha$ and $\psi - \alpha$, and it is assumed that each component is formed of contributions from the two deflectors nearest in direction to its angular position. For example, in Figure 7, $\bar{X}_c$ falls between A and B so that $\bar{X}_c = \bar{A}_c + \bar{B}_c$.

Thus

$$\bar{X} = \bar{X}_c + \bar{X}_a,$$

$$| X_c | = | X_a | = | X |/2 \cos \alpha, \qquad \dots \dots \dots \dots (1)$$

and the contribution of deflector A is

$$| A_c | = \frac{| X |}{2 \cos \alpha \sin (2\pi/5)} \sin \left(\frac{2\pi}{5} - | \psi + \alpha | \right). \qquad \dots \dots (2)$$

As the collector currents may become zero, but may not change in sign, negative values of the sine term in (2) are replaced by zero.

When the clockwise component $\overline{X}_c$ is resolved in this way, for $\alpha = 27°$, the required current waveforms at the front group of collectors appear as shown
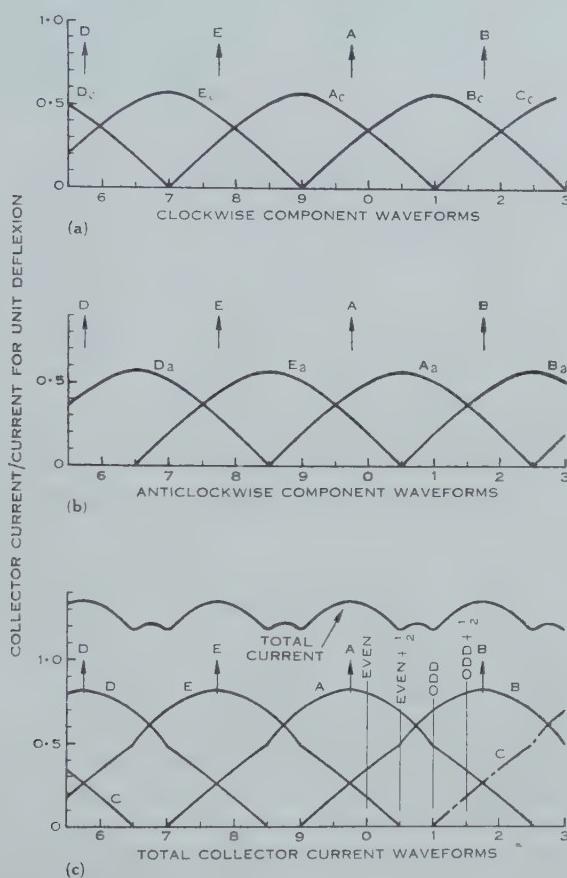
Fig. 8.—Calculated waveform of the rotating field components shown in Figure 7 for $\alpha = 27°$.

  (a) The clockwise component.
  (b) The anticlockwise component.
  (c) The total deflexion.

in Figure 8 (a). The waveforms of the anticlockwise component (Fig. 8 (b)) are found in a similar way. The total deflexion by A is therefore $\overline{A} = \overline{A}_c + \overline{A}_a$ and

$$|A| = \frac{|X|}{2\cos\alpha\sin(2\pi/5)}\left[\sin\left(\frac{2\pi}{5} - |\psi + \alpha|\right) + \sin\left(\frac{2\pi}{5} - |\psi - \alpha|\right)\right]. \quad \cdots \quad (3)$$

The waveforms calculated from (3) for $\alpha = 27°$, shown in Figure 8 (c), are used as a basis for the counter design, although some compromise must be made to simplify the collector shapes.

The stable positions are spaced at approximately 18° intervals symmetrically about the directions of the deflector plates as shown in Figure 8. In 10 of the positions it is essential to divide the beam current between three collectors, as in position 0 (Fig. 8 (c)), and this is done by allowing some secondary electrons to pass from the back collector plate A to E. The secondary current is controlled by the field provided in the design of the collector assembly, and in similar tubes the distribution is quite repeatable. The collector shape seen above each of the odd states in Figure 4 is used also to transfer current to the following plate.
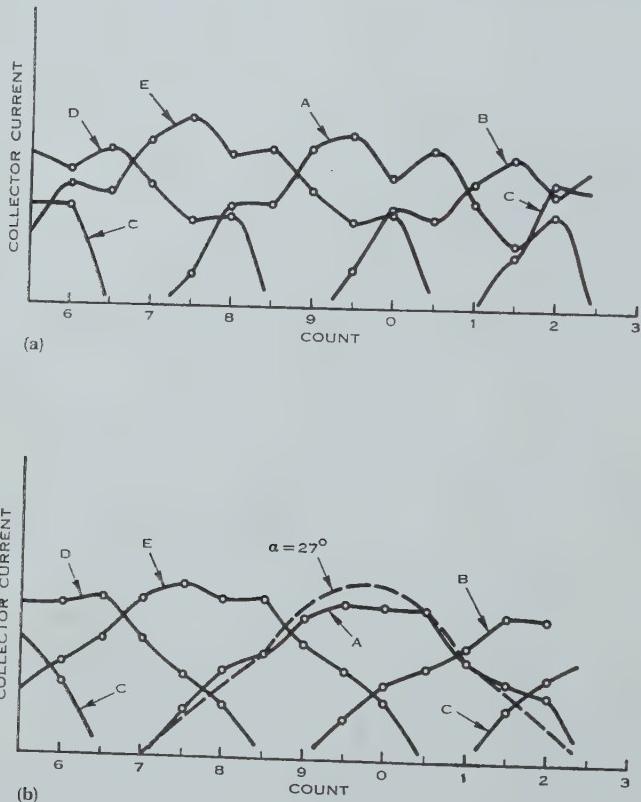
Fig. 9.—Waveform of the collector currents measured on experimental counters.

  (a) An unstable form.
  (b) A stable form.

The two examples of measured collector current waveforms drawn in Figure 9 (a) and (b) show the effect of a change in the division of the back collector current at the even states. In (a), for example at position 0, the current to E is too great in comparison with A, and the stability is improved when this is corrected as shown in (b).

The deflector designs considered in the following section are chiefly those suited to the waveforms given by equation (3), α=27°, and in Figure 8 (c).

III. The Electrode Structure

(a) *The Deflexion Field*

If an initially well-focused electron beam is deflected by reducing the potential of one or two of the five plates of the simple deflector system shown in Figure 1, the circular spot is broadened into an arc. The distorted spot is narrow in the radial direction, but the more important circumferential dimension, which must not exceed an 18° arc, cannot be reduced to the original size by the adjustment of the potential of the focusing electrode. A deflexion field causing this distortion may be termed " divergent ". The opposite form of distortion is shown when the beam is deflected by lowering four plates of the same system equally in potential, or by raising one ; the spot then appears as a radial line, which it is comparatively easy to compress into a spot after deflexion. Therefore it is desirable that the deflexion fields should be either convergent or distortion-less, rather than divergent. For this reason the simple system of five straight plates is not suited to the waveforms of Figure 8 in which groups of two or three plates contribute to the field.

A more uniform field is created by plates extending around an arc greater than $2\pi/5$, and since each plate may occupy only one-fifth of the total surface area of the deflecting cylinder, the effect of larger angles is obtained by twisting the plates into a helical or spiral form.*

It was first confirmed experimentally that a degree of spiralling is effective in reducing the distortion. The changes are most readily shown, however, by the method described in Appendix II, in which the spiralled plates are replaced by an equivalent cylinder varying in potential around the circumference, but constant in the axial direction. The fields are then calculated or measured from models in an electrolytic tank. The values of deflexion field shown in Figures 10 and 11 were determined in this way, and show the effect of increasing the twist of the deflecting plates from zero to one full turn around the deflecting cylinder. The twist of the plates is given as a fraction of one turn by the first figure of the designation, for example 0·6 in 0·6–2. Because the shape of the deflecting field varies cyclically with the angle of deflexion of the spot, ψ, two curves are drawn for each value of twist. These correspond to the two extremes of the variation, one of which occurs when the spot is in line with the direction of a deflector plate, and the second when it is midway between the directions of two adjacent plates. The plate potentials are shown also in Figure 10 ; these are based upon the waveforms of Figure 8 (c).

The curves of field gradient in the central plane (Fig. 10) show that, as the twist of the spiral is increased from zero in the topmost curves, the field becomes more uniform, although lower in intensity, until the spiral extends half way around the circumference, as in the curves 0·5–2, 0·5–3. At larger angles the

---

* A second method of reducing the distortion has also been used, in which 10 or a higher multiple of five straight plates are spaced equally around the deflecting cylinder and connected together to form five similar groups. The central field of a system of 10 plates connected 1–4, 3–6, 5–8, 7–10, 9–2 is convergent, but the deflector diameter must be at least three times larger than the beam, when used with the waveforms of Figure 8 (c).
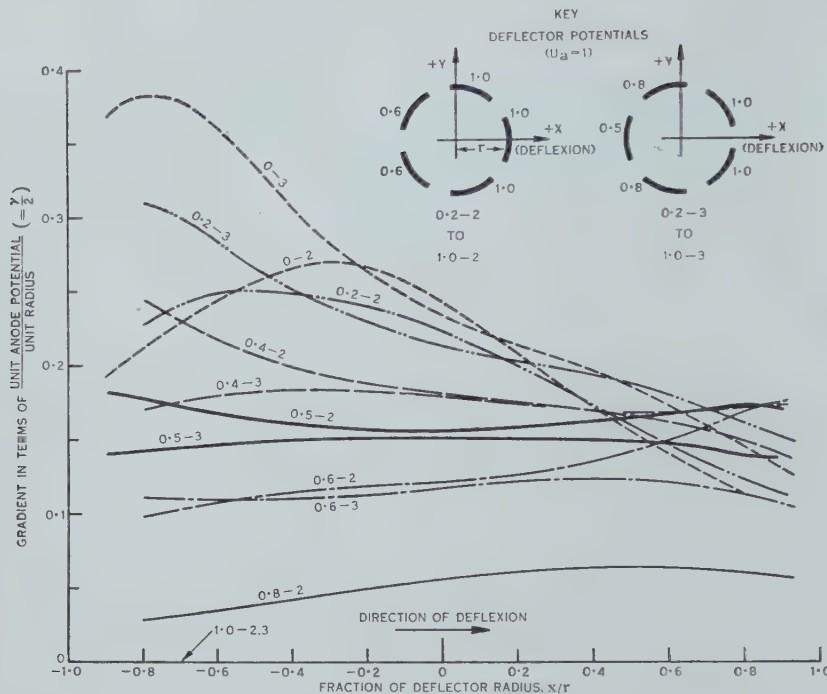
Fig. 10.—The deflexion fields of spiral deflectors.
The most uniform fields are those of curves $0 \cdot 5$–2, and $0 \cdot 5$–3.   In this designa-
tion $0 \cdot 5$ indicates that the spiral electrode makes one-half of one complete
rotation about the deflecting cylinder, and 2, 3, refer to the number of deflectors
lowered in potential together as shown in the key.
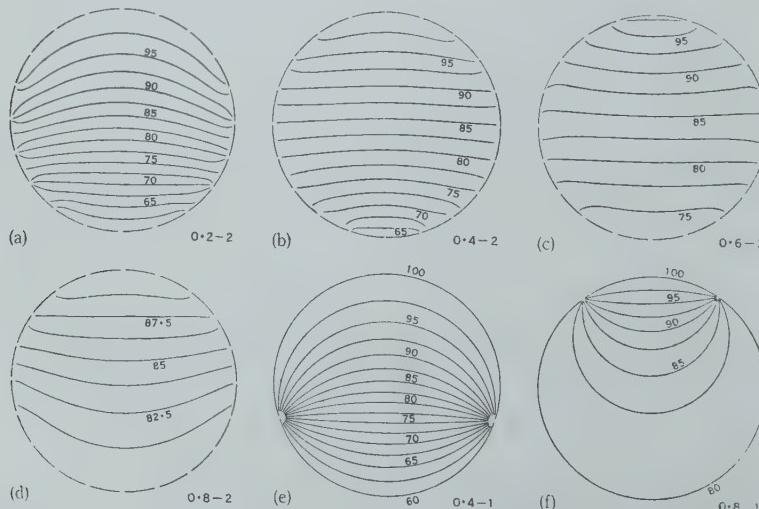
Fig. 11.—Equipotential patterns of the deflexion fields.
The direction of deflexion is upwards in each field, and the designations
correspond with those of Figure 10.

deflexion sensitivity falls more rapidly but the central field becomes slightly convergent, indicated by the upwards slope of the gradient curves in the positive direction, the direction of beam deflexion.

The same changes are shown over the whole area of the potential patterns in Figures 11 (a)-(d). The field becomes more uniform as the twist increases from 0·2 to 0·6, and finally convergent at 0·8, shown by the curvature (concave upwards) of the equipotentials. In this sequence the fall in deflexion sensitivity is indicated by an increase in the spacing of the contours. Although these equipotentials apply to a spot position midway between the directions of two deflectors, similar changes appear also in the group measured for a position directly in line with a deflector when the spot is controlled by three adjacent plates.

These patterns may be compared with the fields of single plates occupying 0·4 and 0·8 of one rotation, shown in Figure 11 (e) and (f). The strong convergence of the single 0·8 spiral, which could be used to maintain a sharply focused spot at a higher beam current, is always lost when two or three plates contribute to the field. Spiral deflectors, twisted in the range 0·3-0·5 of one rotation and including the straight sections shown in Figure 1, have been used in recent counters as a compromise between deflexion sensitivity and distortion.

### (b) *The Electrode Proportions and the Highest Counting Frequency*

The problem of selecting the major dimensions of the electrode system to attain the highest counting frequency will now be considered. As in other electron beam tubes the performance depends upon a large number of variables including the beam current, the anode potential, the change in deflector potential during counting, the spot radius, and all the main electrode dimensions. Several of these are made interdependent by the operating principle and some limits are set by difficulties in construction, but the dominating requirement is the need to maintain maximum stability in all states.

It may be assumed that if the operating principle remains unchanged, the highest counting frequency of a scale-of-N tube is proportional to

$$\frac{I}{C(U_a - U_d)/N} \quad \cdots\cdots\cdots\cdots\cdots\cdots (4)$$

where C = an effective value of capacitance associated with one of the deflector-collector combinations,

       I = beam current,

$U_a - U_d$ = the difference between the highest and lowest potentials reached by a deflector during counting. $(U_a - U_d)/N$ is a measure of the deflector potential change taking place between one state and the next.

At this point it is convenient to divide the problem into two parts, and to consider firstly the best size of the tube as a whole, and secondly the best proportions or shape of the electrode system.

The first question is answered by the scaling relation termed by Moss (1945) " The Principle of Geometrical Similitude ". This principle states that if all the dimensions of an electrode system are multiplied by some constant factor, the

operating potentials and beam current remain unchanged. However, the capacitance C will vary linearly in proportion to the scale factor, and therefore the counter as a whole should be reduced in size as far as the methods of construction allow.

In the second part of the problem it will be assumed that the overall size is chosen so that the capacitance C remains constant. Apart from its simplicity, this assumption of constant capacitance is more in accordance with a limitation set by the difficulty of constructing very small components than the alternative of keeping any one dimension constant. Thus the electrodes must be proportioned to make the term $NI/(U_a - U_d)$ a maximum, and for convenience in analysis, the dimensions will be found in terms of the initial radius of the electron beam cross section.

The allowable spot radius s depends upon the radial deflexion of the beam at the collectors. For stability, s may not exceed the value derived in Appendix I, and given by

$$l = s^{0.65} d^{0.35} U_a^{0.52} (4N/\pi)^{0.65} / \gamma^{0.5} (U_a - U_d)^{0.5}. \quad \ldots \ldots \ldots (5)$$

It has been shown (Hollway 1952) that the greatest value of beam current I, in milliamperes, which can be concentrated into a spot of radius s, is governed by the length of the beam and the potential (in volts) where

$$l = 0.39 s^{0.4} U_a^{0.75} / I^{0.5}. \quad \ldots \ldots \ldots \ldots \ldots (6)$$

Therefore, on combining equations (5) and (6),

$$\frac{NI}{(U_a - U_d)} = 0.028 U_a^{0.5} \gamma / d^{0.7} s^{0.5} N^{0.3}. \quad \ldots \ldots \ldots \ldots (7)$$

Thus, when the capacitance C is constant, the frequency varies as $s^{-0.5}$, so that the spot radius should be kept small. This is explained by the fact that, if the length of the structure is increased, the allowable spot radius increases (with the deflexion distance) in proportion to $l^{1.54}$; but the actual spot radius, controlled by space-charge defocusing, rises more rapidly, as $l^{2.5}$, so that a small value of l, and of s, is needed.

The possibility of raising the sensitivity (γ) and improving the uniformity of the deflexion field, which sets a lower limit to d, has been discussed in Section III (a).

When equation (5) is compared with results measured on experimental counters, close agreement is found. However, the beam current predicted by equation (6), assuming beam potentials of U_a and U_d, is 4 and 1.5 times the observed maximum because spherical aberration in the objective lens, deflexion defocusing, and the geometrical spot size cannot be made negligible.

IV. Performance

The electrical operating characteristics of the counter are contained in Figures 12 and 13. The anode characteristic (Fig. 12) is based upon the current flowing to the positive ring. During counting the beam current must be sufficient to hold the beam against the edge of the positive ring so that it is partially intercepted. At the $9\frac{1}{2}$ state, the ring current is transferred to the carry electrode

(*R*, Fig. 2) where it may be measured as an indication of the correct grid bias at any anode potential. Figure 12 shows the change in carry current with anode voltage, at different values of cathode resistance. The upper limit to the carry current is set by the increase in spot size which takes place when the beam current is raised. In the operating range the anode potential is not critical and no stabilization is necessary.
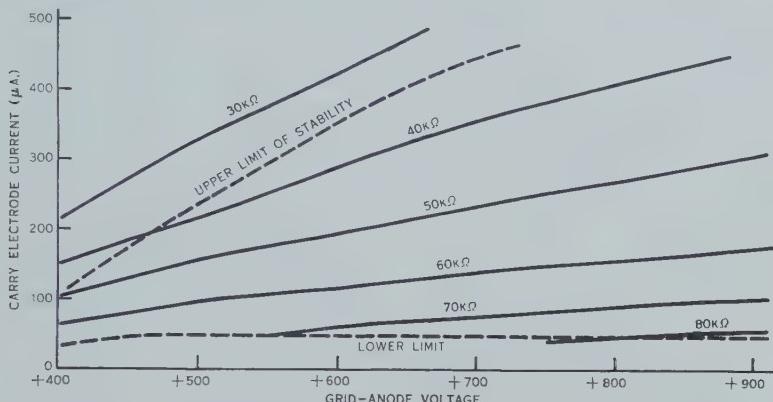
Fig. 12.—Anode potential characteristic.
The curves show stable operating ranges for different values of cathode
(self-bias) resistance.

The potential levels at which changes occur between states and the ranges of stability are shown in the trigger characteristic of Figure 13. The extent of the range in the negative direction may be explained by the rapid decrease in the trigger sensitivity as the beam moves inwards. In the positive direction
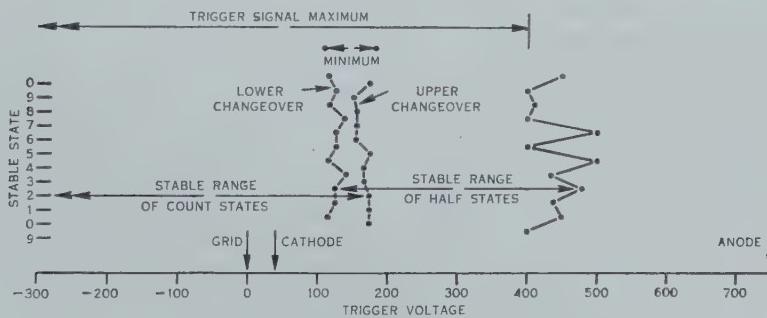
Fig. 13.—Trigger characteristic.
The measured points define ranges of the trigger potential in which the count
and intermediate positions of the beam are stable.

the range ends when the trigger fails to suppress secondary electrons from the front plates. The power supply and coupling circuits may be designed from these characteristics.

The complete circuit of a counter having four scale-of-10 stages is shown in Figure 14. Four of the five collector resistances in each stage are returned to the anode line, which is held at approximately 750 V. ; the resistances connected

to the 0–1 plates are returned to the same potential through the reset push button.

When a negative-going signal appears at the input terminals, the grid of the first triode section falls, raising the trigger of the first counter and bringing the beam into an intermediate position. The supply potential of the triode anode resistors is chosen within the stable range of the intermediate states, so that the incoming signals are not restricted in amplitude. The count is completed when the trigger potential falls to the lower change-over point. The usual pulse forming or " trigger circuit ", needed with multiple tube counters operating from zero frequency upwards, is unnecessary because discontinuities result from the form of the collector plates. Also, because the speed at which the beam
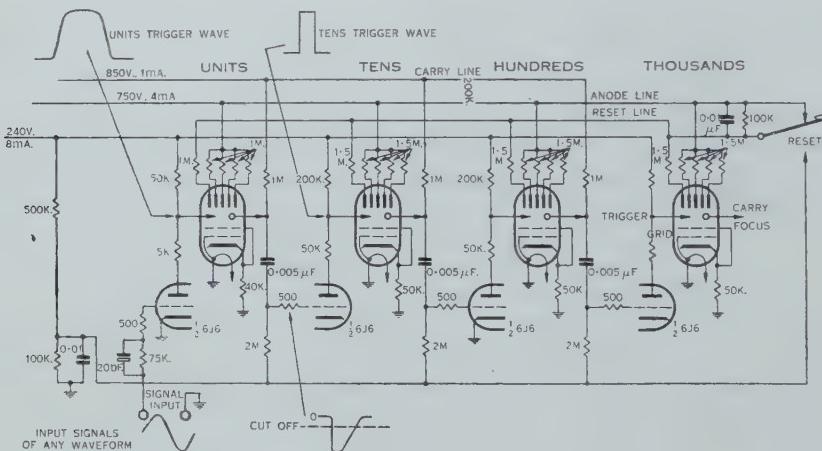
Fig. 14.—Circuit of a complete four-tube counter.

rotates is independent of the rate of change of the trigger potential, the pulse generated at the carry electrode when the beam falls into the $9\frac{1}{2}$ position is always sharp and may be coupled by means of a capacitance to the grid of the second triode, without thereby setting a lower limit to the speed of counting.

Between pulses the triode sections draw grid currents of 20–50μA. through their grid resistances. This eliminates stray signals below a certain amplitude, for example those appearing at the carry electrode before the $9\frac{1}{2}$ state is reached.

In resetting the counter, the line connected to the resistors of the 0–1 collectors is lowered momentarily in potential, bringing all the electron beams to the zero positions. Carry-over pulses are often produced during resetting and these are suppressed by increasing the grid currents of the triodes through the additional circuit from the reset push button shown in Figure 14.

At frequencies approaching the highest speed of counting, there is a fall in the amplitude of the carry signals caused by failure of the deflectors to return to the full anode potential in the intervals between transits of the beam. Beyond the upper limit the beam fails to complete one rotation during 10 incoming signals and the divisor rises to values higher than 10. This limit has been measured at frequencies between 70 and 180 kc/s., depending upon the circuit conditions.

V. Future Developments

The basic principle of the " self-rotating " beam may be applied to a number of tubes for special purposes. One of these is a reversible counter which will add or subtract signals from the stored count. For subtraction, a third ring of intermediate stable states can be used, on a smaller radius than the count positions ; change-over on the inner path is in the anticlockwise direction, so that each negative trigger signal subtracts one unit from the count.

Either the reversible form or the normal counter tube may be constructed with two or more trigger rings instead of one ; this would allow signals from two or more sources to add to or subtract from the count, while maintaining electrical isolation between the inputs. If one trigger receives only positive signals and the second is restricted to negative ones, coincident signals will cancel and produce no change in the count.

By a modification to the collector system, the counter may be extended to accept without error coincident positive signals from a pair of triggers, combinations of positive and negative, and coincident negative signals. This is accomplished by the use of an additional ring of 10 stable states, in a scale-of-10 tube, on a larger radius than the half states ($1\frac{1}{2}$, $2\frac{1}{2}$, $3\frac{1}{2}$, etc.). States in the outer ring lie midway between the half states, and are therefore in the same angular positions as the count states (1, 2, 3, etc.). The outer states may be identified as 1', 2', 3', etc. in correspondence with those in the inner ring. On this collector system the simultaneous arrival of two positive signals moves the spot, originally at the count of 3, for example, to $3\frac{1}{2}$ and beyond this to 4' in the outermost ring. When the signals decrease, the spot moves on first to $4\frac{1}{2}$, and finally returns to 5, which is an increase in the stored count of two units. Thus to count accurately coincident signals from a number of input channels, the number of complete rings of states needed in addition to the " count " states is equal to the number of channels. The minimum value of the trigger signal rise time, which is not restricted in the simple counter, must be large enough in the coincidence form to allow the necessary angular rotation of the spot. Coincidence tubes, used in the product register of a decimal electronic computer, could accept carry-over signals occurring during arithmetic operations.

Scale-of-N counter tubes may be constructed on the same principle, either by changing the number of stable states in each collector or by changing the number of deflectors and collectors. For example a scale-of-12 may use six deflectors and collectors, each collector having four states, as in the scale-of-10. For a scale-of-5, five deflectors and collectors having two stable states in each collector would be more suitable.

To provide separate output signals corresponding to each digit, for some applications it may be necessary to construct a scale-of-10 counter with 10 plates instead of five, and it would be simpler to use untwisted deflectors raised in potential by secondary emission. This would allow the use of strongly convergent fields (see Section III (*a*)), either for the deflexion of a beam, or to create localized gradients at a cathode surface to direct and focus the electron stream without the need for an electron gun or deflector system.

H

VI. Acknowledgments

VII. References

Gabor, D. (1937).—*Nature* **139** : 373.
Hollway, D. L. (1950).—*Nature* **165** : 856.
Hollway, D. L. (1952).—*Aust. J. Sci. Res.* A **5** : 430.
Moss, H. (1945).—*J. Brit. Instn. Radio Engrs.* **5** : 204.
Salinger, H. (1937).—*Electronics* **10** : 50.
Snyder, R. L. (1946).—U.S. Pat. 2,404,106.
Wynn-Williams, C. E. (1931).—*Proc. Roy. Soc.* A **132** : 295.

Appendix I

The Allowable Size of the Deflected Spot

In a scale-of-N counter each one of the $2N$ stable states consists of two regions, as described in Section II (a). It will be assumed that the electron spot, of radius s, must fall completely within one of these regions in order to develop fully the stable margins at each state. Thus the necessary deflexion x is given by

$$x = 4Ns/\pi. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (8)$$

An expression for the deflexion will now be derived from the electrode geometry.

To improve the deflexion sensitivity practical deflectors are bellmouthed (e.g. Fig. 2) and it is desirable to include this shaping in the derivation. It is assumed therefore that the inside radius of the deflector surface follows the central trajectory of the deflected beam but at a distance d from it. As in Section III (b), the initial radius of the electron beam is taken as the unit of length, and the quantities d, l, x, and s, represent numbers of beam radii.

The motion of an electron entering the deflector along the central axis is given by

$$\mathrm{d}^2x/\mathrm{d}t^2 = e\gamma(U_a - U_d)/m(x+d) \text{ (e.s.u.)}, \quad \ldots\ldots\ldots\ldots (9)$$

where γ is constant for a particular deflector design and waveform, relating the intensity of the deflexion field to $(U_a - U_d)/(x+d)$. (Values of γ for a number of different deflectors are shown in Figure 10.)

Multiplying both sides of equation (9) by $\mathrm{d}x/\mathrm{d}t$, integrating and inserting the initial conditions $x=0$, $\mathrm{d}x/\mathrm{d}t=0$, when $t=0$:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \sqrt{\frac{2e}{m}\gamma(U_a - U_d)\ln\left(\frac{x}{d}+1\right)} \text{ (e.s.u.).} \quad \ldots\ldots\ldots (10)$$

Since the length l is measured along the beam (from $l=0$, $x=0$) and the beam potential is near U_a throughout the trajectory, as shown in Figure 11, $\sqrt{2e/m\,U_a}$ (e.s.u.) may be substituted for $\mathrm{d}l/\mathrm{d}t$, giving

$$\frac{\mathrm{d}x}{\mathrm{d}l} = \sqrt{\gamma(1 - U_d/U_a)\ln\left(\frac{x}{d}+1\right)}. \quad \ldots\ldots\ldots\ldots (11)$$

On substituting $\ln (x/d+1)=g^2$,

$$l=\frac{2d}{\sqrt{\gamma(1-U_d/U_a)}}\int_0^{\sqrt{\ln (x/d+1)}} e^{g^2}dg. \qquad \dots\dots\dots\dots (12)$$

This form of integral appears in analyses of space charge defocused beams, and cannot be solved in terms of usual functions. However, in the present application, (12) is used to find the final deflexion x of the beam after travelling a distance l, so that values of x/d outside the range $1<x/d<10$ are not needed in practice. With this restriction, a simple substitution is possible and

$$l\approx 2d(x/d)^{0.65}/\gamma^{0.5}(1-U_a/U_d)^{0.5}. \qquad \dots\dots\dots\dots\dots (13)$$

Therefore combining equations (8) and (13), the greatest allowable spot radius s is given by the relation

$$l\approx s^{0.65}d^{0.35}U_a{}^{0.52}(4N/\pi)^{0.65}/\gamma^{0.5}(U_a-U_d)^{0.5}. \qquad \dots\dots\dots (5)$$

APPENDIX II

The Determination of the Deflexion Fields and Electron Paths

The circle $AHCD$, of radius r (Fig. 15) represents a cross section through a long deflector system of which the dotted segment DAH is at zero potential and the remainder of the circumference is at the potential U_0. To convert the electric field pattern of the deflector to a simple form, the section is transformed conformally by inversion with the point H as centre, so that any point M having coordinates x, y in the original section is represented by M' in accordance with the transforming equation,

$$HM \cdot HM'=(2r)^2. \qquad \dots\dots\dots\dots\dots (14)$$

The dotted segment DAH becomes the straight line $D'A'H'$ at zero potential, and $D'C'H'$ represents the remainder of the circumference at potential U_0; the X-axis is converted to the semicircle $A'C'$ having D' as centre.

In the half-plane P, lying to the right of $H'D'H'$, which corresponds to the interior of the circle, the equipotentials are all straight lines radiating from D', so that the potential U_{xy} of point M is the potential of M', or

$$U_{xy}=U_0\beta/\pi.$$

From the geometry of the figure, β may be found in terms of x, y, and θ, giving

$$U_{xy}=\frac{U_0}{\pi}\left[\theta+\cot^{-1}\tfrac{1}{2}\left\{\frac{\sin \theta}{x/r+\cos \theta}\left(1-\frac{(y/r)^2}{\sin^2 \theta}\right)-\frac{x/r+\cos \theta}{\sin \theta}\right\}\right]. \qquad \dots\dots (15)$$

The potential along the X-axis may be found by setting $y=0$ in (15), or directly from the figure, since $HD'=D'G'=D'E'$, $\angle G'D'E'=2\varphi$, and therefore

$$U_x=U_0(\theta+2\varphi)/\pi=\frac{U_0}{\pi}\left[\theta+2\tan^{-1}\left\{\frac{x/r+\cos \theta}{\sin \theta}\right\}\right]. \qquad \dots\dots (16)$$

Because the equipotentials in the P-plane are straight lines through D', they form circles in the XY-plane passing through H and D. The radius of an equipotential U_x which crosses the X-axis at x, is therefore

$$R_x = r\left(\frac{x/r + \cos\theta}{2} + \frac{\sin^2\theta}{x/r + \cos\theta}\right)$$

and, by substitution from (16),

$$R_x = r\sin\theta / 4\sin\left(\frac{\pi U_x}{U} - \theta\right). \quad \cdots\cdots\cdots\cdots (17)$$
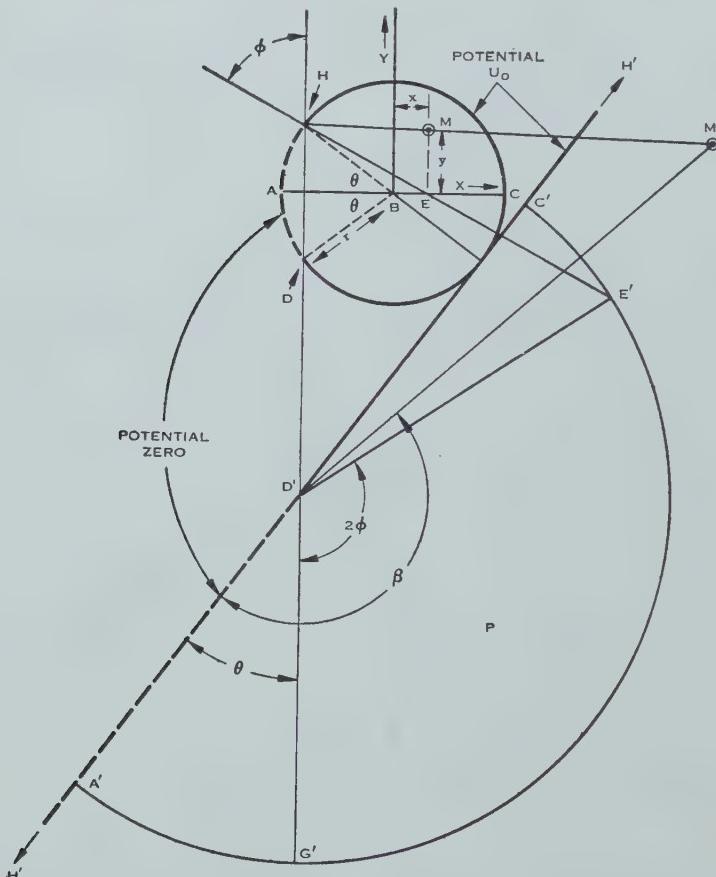
Fig. 15

The gradient along the X-axis is given by the differentiation of (16),

$$\frac{\partial U_x}{\partial x} = \frac{U}{\pi r} \frac{2\sin\theta}{(x^2/r^2 + 2x/r\cos\theta + 1)}. \quad \cdots\cdots\cdots\cdots (18)$$

The maximum value of the gradient occurs at $x/r = -\cos\theta$, that is, on the chord joining the ends of the segment, which is also an equipotential of infinite radius. Thus as θ is increased to approach $\pi/2$, the position of the maximum

moves from negative values of x/r towards the centre of the deflexion field where it coincides broadly with the position of the electron beam cross section. This more favourable position of the maximum explains the comparatively small decrease in the central gradient shown by the sequence of the topmost six curves in Figure 10 and the subsequent rapid drop in sensitivity when the twist is increased beyond $\theta/\pi = 0 \cdot 5$.

Equation (18) has been used to find the deflexion field intensity of spiralled deflectors by the following approximate analysis. First the potentials of the five plates, corresponding to a chosen position of the spot, are found as described in Section II (b). The spiralled deflector system is then converted to one similar in general form but composed of untwisted strips. Each strip (of angular width θ_1) is given a potential U'_θ equal to the mean potential of the corresponding surface area of the spiralled system. Thus, for a cylindrical array of length Z_d,

$$U'_\theta = \frac{1}{\theta_1 Z_d} \int_0^{Z_d} \int_{\theta-\theta_1/2}^{\theta+\theta_1/2} U\,d\theta\,dz. \quad \ldots\ldots\ldots\ldots\ldots (19)$$

It is convenient to set $\theta_1 = 18°$ and to integrate graphically. The gradient is then calculated from these values of U'_θ by superposition in equation (18). However, it is usually quicker to set up the potentials U'_θ in an electrolytic tank model of the untwisted system, and to measure the gradient and potential directly. The accuracy of the tank measurements, judged from a comparison between measured and calculated curves of Figure 10, was approximately ± 3 per cent. of $\partial U/\partial x$.

To find the electron paths in these deflexion fields, which are non-uniform but symmetrical about the X-axis, the " circle method " (Salinger 1937) has been used in both the XY- and XZ-planes. Recently an electron tracer based on Gabor's Tangent Bridge (Gabor 1937) was constructed and found to be quicker and more accurate. However, in problems of figure projection, the distribution of beam current between the collectors, constructional tolerances and others, it is necessary to supplement electron path tracings with measurements made on trial electrode assemblies.

VALIDITY OF MATTHIESSEN'S RULE FOR COLD–WORKED WIRES

By W. Boas* and J. F. Nicholas*

Summary

Matthiessen's rule has been tested by determining the slopes of the electrical resistivity-temperature curves for wires of eight common metals and alloys in various states of deformation by wire-drawing. The results show that the slope is independent of deformation, i.e. the rule is true, to within $0 \cdot 5$ per cent. for nickel, copper (two purities), iron, and 80/20 brass, and to within 1 per cent. for aluminium. However, for 75/25 brass and an aluminium bronze, deformations corresponding to logarithmic strains of $2 \cdot 3$ decrease the slopes by 1 and 3 per cent. respectively. As an explanation of this behaviour, it is suggested that deformation causes an increase in the characteristic temperature.

I. Introduction

Matthiessen's rule states that, for a metal or alloy, the product of the electrical resistivity and its temperature coefficient is not affected by either deformation or the addition of small concentrations of solute atoms. Whereas the effect of alloying has been investigated extensively, that of deformation has not attracted much attention. During recent work in this laboratory (Broom 1952) on the resistivity of cold-drawn wires, some results were found which suggested deviations from the rule. The most recent systematic work on the validity of the rule for a metal in various states of deformation is that by Geiss and van Liempt (1925, 1927) who found that, for tungsten, molybdenum, platinum, and nickel, deformation produced variations in the product of up to 2 per cent. They give no indication of their experimental error but conclude that the rule is true. As the rule implies the possibility of separating the resistivity of a metal into a part dependent only on temperature and a part dependent only on deformation, it is of interest to know how closely it is obeyed. Therefore, an investigation has been carried out to test its validity for wires of eight common metals and alloys, the accuracy aimed at being $0 \cdot 5$ per cent.

Since, by definition, the temperature coefficient of resistivity, α, is given by

$$\alpha = \frac{1}{\rho} \frac{d\rho}{dT},$$

where ρ is the resistivity and T the absolute temperature, Matthiessen's rule can be formulated as either $\alpha\rho = $ constant or $d\rho/dT = $ constant. In this investigation the rule was tested by considering the variation of $d\rho/dT$ with deformation.

* Division of Tribophysics, C.S.I.R.O., University of Melbourne.

II. Experimental

(a) *Preparation of Specimens*

The materials used and their compositions are given in Table 1.

TABLE 1

COMPOSITIONS OF MATERIALS USED

Nickel	0·03% C; 0·01% Co; 0·12% Mg; <0·05% Mn, Cr, Cu
Spectrographic copper	0·0003% Ag, Ni; <0·0004% Pb
Conductivity copper ..	<0·01% Ag, Cd, Pb, Si; ≪0·01% Mg, Fe
Spectrographic iron ..	0·002% Cu; 0·005% Ni
Spectrographic aluminium	0·002% Mg; <0·001% Si; <0·0005% Fe, Cu
80/20 Brass	20·21% Zn; 0·02% Fe; 0·01% Pb, Sn
75/25 Brass	25·35% Zn; 0·1% Fe; 0·05% Sn; <0·01% Pb; ≪0·01% Mn
Aluminium bronze ..	5·67% Al; 0·28% Zn; 0·1% Fe

These were originally in the form of annealed wires of approximately 4 mm. diameter and were then drawn at room temperature to a final diameter of approximately 1 mm. Each time the wire was passed through a die, a length was cut off and annealed for 2 hr. at 600 °C. (400 °C. for aluminium) before the original wire and the cut-off pieces were deformed further. In this way, 11 specimens of each material were obtained, all of similar dimensions (30 cm. long and 1 mm. diameter) but each having been deformed by a different amount since the last annealing. The deformations corresponded to logarithmic strains spread evenly over the range 0–2·3. The similarity in specimen sizes meant that, for each material, all measurements of the same quantity and the corresponding experimental errors were of the same magnitude.

(b) *Measurements*

Using a Kelvin double bridge, the resistance of a standard length of each specimen was measured in a bath of melting ice (0 °C.) and in one of liquid oxygen (−183 °C.), all measurements being repeated two or three times. The weights and lengths of the specimens were then measured and, from these, relative values of the mean cross-sectional areas for all specimens of the one material were calculated. Finally, the specimens were annealed (as above) and the resistance measurements repeated on the annealed wires. A few check experiments showed that the changes in length and weight which occurred on annealing were negligible and it was assumed that the density remained constant, so that the geometry of the specimen was taken to be unaltered on annealing.

The temperature of the ice-bath was measured to within 0·1 °C. with a calibrated mercury-in-glass thermometer graduated to this accuracy. When resistance measurements were made the temperature of this bath was usually between +0·5 and −0·5 °C. The temperature of the liquid oxygen bath was measured at intervals with a pentane-in-glass thermometer and the reading of the thermometer was always within ±0·1 °C. of −183·2 °C. A continuous

check on the constancy was maintained by means of a thermocouple. The calibration of the thermometer was not checked because any error in its reading would have affected only the absolute values of $d\rho/dT$ and not the relative ones, since the variations in temperature of the baths were small compared with the total temperature difference.

III. RESULTS

The actual results are expressed in terms of $\Delta\rho/\Delta T$, where Δ denotes the change in value between measurements made in the two baths. The subscripts D and A will be used to denote deformed and annealed values respectively.

TABLE 2

SUMMARY OF RESULTS

1	2	3	4	5	6	7
Metal	$\Delta\rho/\Delta T$ (Normalized Values)		Maximum Deviations of $(\Delta\rho/\Delta T)_D$ from Mean of Annealed Values		$\left(\dfrac{(\Delta R/\Delta T)_D}{(\Delta R/\Delta T)_A}-1\right)\times 100$	Increase in ρ 0°C. after Logarithmic Strain of $2\cdot 3$ (%)
	Annealed	Deformed	+ve	—ve		
Nickel	$100\pm 0\cdot 25$	$100\cdot 11\pm 0\cdot 22$	$0\cdot 45$	$0\cdot 23$	$0\cdot 16\pm 0\cdot 13$	6
Spectrographic copper	$100\pm 0\cdot 18$	$100\cdot 01\pm 0\cdot 15$	$0\cdot 29$	$0\cdot 26$	$-0\cdot 06\pm 0\cdot 26$	3
Conductivity copper	$100\pm 0\cdot 24$	$99\cdot 99\pm 0\cdot 15$	$0\cdot 12$	$0\cdot 34$	$0\cdot 02\pm 0\cdot 30$	3
Spectrographic iron	$100\pm 0\cdot 19$	$100\cdot 09\pm 0\cdot 22$	$0\cdot 36$	$0\cdot 34$	$-0\cdot 03\pm 0\cdot 42$	1
Spectrographic aluminium	$100\pm 0\cdot 38$	$99\cdot 92\pm 0\cdot 54$	$0\cdot 98$	$0\cdot 74$	$0\cdot 00\pm 0\cdot 50$	$0\cdot 5$
80/20 Brass	$100\pm 0\cdot 24$	$100\cdot 10\pm 0\cdot 15$	$0\cdot 38$	$0\cdot 13$	$-0\cdot 06\pm 0\cdot 66$	20
75/25 Brass	$100\pm 0\cdot 30$	$99\cdot 47\pm 0\cdot 50$	$0\cdot 10$	$1\cdot 00$	$-0\cdot 35\pm 0\cdot 43$	25
Aluminium bronze	$100\pm 0\cdot 32$	$98\cdot 01\pm 1\cdot 46$	—	$3\cdot 88$	$-1\cdot 98\pm 1\cdot 69$	28

Absolute values of $\Delta\rho/\Delta T$ would have involved knowledge of the densities of the specimens and of the distance between the potential knife-edges in the resistance bridge as well as an accurate calibration of the low temperature thermometer. Therefore, only relative values are given in the results shown in Table 2, a normalizing factor having been applied to make the mean value of $(\Delta\rho/\Delta T)_A$ equal to 100 in each case. This procedure facilitates comparison of the standard deviations.

In Table 2 the second column also gives the standard deviation of $(\Delta\rho/\Delta T)_A$, the third gives the mean and standard deviation of $(\Delta\rho/\Delta T)_D$, while the fourth and fifth give the maximum deviation of $(\Delta\rho/\Delta T)_D$ from the normalized mean

of $(\Delta\rho/\Delta T)_A$. It should be noted that the values in the third column are meaningless if Matthiessen's rule is not true, i.e. if $\Delta\rho/\Delta T$ does vary with deformation, since the individual results from which they are derived refer to different amounts of deformation (logarithmic strains between $0\cdot2$ and $2\cdot3$). However, if the rule is true, then they have a meaning and should agree with the values in the second column.

For the first six metals there is close agreement between the values in the second and third columns and, furthermore, the range of values of $(\Delta\rho/\Delta T)_D$ is only of the magnitude to be expected from the experimental errors, so that it can be concluded that Matthiessen's rule is true within an accuracy of $0\cdot5$ per cent. for nickel, copper (two purities), iron, and 80/20 brass, and within 1 per cent. for aluminium.* However, for the 75/25 brass and the aluminium bronze the mean values of $(\Delta\rho/\Delta T)_D$ are below 100, the standard deviations are large and the maximum deviations are large and asymmetric. In Figure 1, the values of
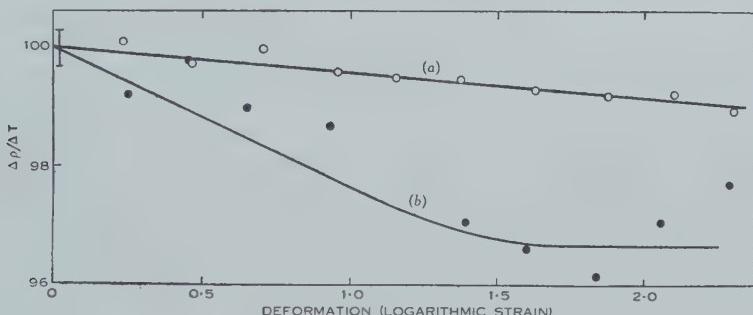

Fig. 1.—Variation of $\Delta\rho/\Delta T$ with deformation for (a) 75/25 brass and (b) aluminium bronze. The vertical bar at the left-hand side shows the standard deviation of the values of $(\Delta\rho/\Delta T)_A$.

$(\Delta\rho/\Delta T)_D$ are plotted against deformation for these two metals and it can be seen that $(\Delta\rho/\Delta T)_D$ decreases as the deformation increases. Therefore, it can be said that these two metals show negative deviations from Matthiessen's rule, the magnitude of the deviations after a logarithmic strain of $2\cdot3$ being approximately 1 per cent. for the brass and 3 per cent. for the aluminium bronze.

Since the geometry of any particular specimen is constant, we have the relation

$$\frac{(\Delta\rho/\Delta T)_D}{(\Delta\rho/\Delta T)_A}=\frac{(\Delta R/\Delta T)_D}{(\Delta R/\Delta T)_A},$$

where R denotes the resistance. This provides an alternative method of testing the rule and one which does not involve the measurements of length and weight. The test is whether these ratios are unity for all degrees of deformation or whether there is a significant deviation from this value. Table 2 gives the mean and

---

* The larger experimental variation for aluminium was probably due to the softness of this material causing difficulties in fixing the wires in the resistance bridge so that they were straight and unstrained while still maintaining good electrical contacts with the potential knife-edges in both the ice and the liquid oxygen baths.

standard deviation of the differences between these ratios and unity, but the same restrictions on interpretation apply here as applied for the values of $(\Delta\rho/\Delta T)_D$. It can be seen that there is complete agreement between the two methods of analysing the results.

Table 2 also gives values for the percentage increase, due to deformation, in the resistivity at 0 °C. Comparison shows that the only materials which do not follow Matthiessen's rule are alloys which have a large increase in resistivity on deformation, but the converse is not true as the 80/20 brass shows a large increase in resistivity and still follows the rule.

IV. Discussion

The main source of errors in resistivity measurements on wires is the determination of the mean cross-sectional area. The use of a micrometer to measure the diameter of fine wires directly can produce serious errors due to compression between the jaws. The method of weight and length measurements was therefore used to give relative values of the cross-sectional area when needed, but the results as expressed in the sixth column of Table 2 are independent of the wire dimensions. In this connection, the results of Rutter and Reekie (1950) should be noted. These workers measured the resistivities of spectrographic copper and aluminium as functions of deformation for temperatures of measurement of 20, 90, and 297 °K. They give no indication of the accuracy of their results nor of the apparatus used to measure the resistivities but their curves for the resistivity of aluminium as a function of deformation are of different shapes from those usually found (see, for example, Broom 1952). It is suggested here that this difference may be due to errors of measurements of specimen dimensions and that the accuracy of the results is probably at most 0·5 per cent. Within this accuracy, their results for aluminium agree with those in the present work, but the results for copper cannot be compared as sufficient data are not given.

The use of $\Delta\rho/\Delta T$ as a measure of $d\rho/dT$ in the present work does not depend on ρ being an exactly linear function of T in this temperature range, as the deviations from linearity are small and the equality of $(\Delta\rho/\Delta T)_D$ and $(\Delta\rho/\Delta T)_A$ therefore still implies parallelism of the two ρ v. T curves. This is, of course, the condition which allows separation of the resistivity into a part depending only on temperature and a part depending only on deformation.

It is of interest to note that, in the two cases where a deviation from Matthiessen's rule has been found, this deviation is such that the increase in resistivity with temperature is smaller in the deformed metal than in the annealed metal. According to theory, the electrical resistivity due to thermal vibrations, ρ_T, at not too low temperatures is given by

$$\rho_T = C\frac{T}{\Theta^2}\left(1 - \frac{\Theta^2}{18T^2}\right),$$

where Θ is the characteristic temperature and C is a constant involving atomic constants and the number of conduction electrons per unit volume. Assuming

that the deformation dependent part is unaffected by temperature changes, it follows that

$$\frac{d\rho}{dT} = \frac{d\rho_T}{dT} = C\left(\frac{1}{\Theta^2} + \frac{1}{18T^2}\right),$$

so that the slope of the ρ v. T curve depends mainly on Θ since Θ and T are of the same order of magnitude. If the number of conduction electrons is not altered by plastic deformation the observed change in $d\rho/dT$ on deformation can then be expressed in terms of an increase in Θ :

$$\delta(d\rho/dT) = -(2C/\Theta^3)\delta\Theta.$$

The fractional increase in $d\rho/dT$ can therefore be expressed as $-2\delta\Theta/\Theta$, so that, for the two alloys being considered, $\delta\Theta$ is positive and of the order of a few degrees.

This, of course, is only another way of expressing the experimental results but it does suggest a possible physical interpretation. Any deviation from Matthiessen's rule implies that the scattering of conduction electrons due to thermal vibrations is not independent of that due to lattice disturbances produced by plastic deformation. This may mean either that the scattering probabilities are not simply additive, or that the thermal vibrations in a distorted crystal are different from those in a perfect crystal. It is here suggested that the latter is the case and that the deviations from Matthiessen's rule indicate that deformation increases the maximum frequency of the lattice vibrations, thereby raising the characteristic temperature. A detailed calculation is required to show whether the inhomogeneous strain fields in a deformed crystal do have such an effect on the lattice vibrations. In this case a corresponding change in the relevant elastic constant would also be expected (Blackman 1951).

V. Acknowledgments

VI. References

Blackman, M. (1951).—*Proc. Phys. Soc. Lond.* A **64** : 681.
Broom, T. (1952).—*Proc. Phys. Soc. Lond.* B **65** : 871.
Geiss, W., and van Liempt, J. A. M. (1925).—*Z. anorg. Chem.* **143** : 259.
Geiss, W., and van Liempt, J. A. M. (1927).—*Z. Phys.* **41** : 867.
Rutter, J. W., and Reekie, J. (1950).—*Phys. Rev.* **78** : 70.

THE FAILURE OF MATTHIESSEN'S RULE FOR HEAVILY DEFORMED ALLOYS

By P. G. KLEMENS*

Summary

Deviations from Matthiessen's rule, such as were found by Boas and Nicholas (1953) for some heavily deformed alloys, may be explained by assuming that in the strongly disordered regions the thermal vibrations of the atoms contribute much less to the resistance than in perfect regions of the crystal. The magnitude of the calculated effect is compatible with the experimental results.

I. GENERAL ASSUMPTIONS

In the preceding paper (Boas and Nicholas 1953) the occurrence of deviations from Matthiessen's rule for some heavily deformed alloys has been discussed. Whereas these authors ascribe the deviations to a change in the frequency of the atoms vibrating in a crystal containing strains due to plastic deformation, it is here suggested that these deviations are caused by a non-linear additivity of the scattering probabilities of the conduction electrons. It is the purpose of this paper to show that the experimental results are compatible with this explanation.

As in the preceding paper it is assumed that the component of resistivity due to elastic scattering by the deformations is unaffected by temperature changes. It is now further assumed that in the neighbourhood of a static imperfection produced by plastic deformation the displacement of the atoms is large compared with their average thermal displacement. Then the scattering probability of the conduction electrons due to thermal displacements in this region will not be proportional to the mean square thermal displacement of the atoms. There will be a saturation effect and the thermal displacement of an atom will contribute much less to the resistance than it would in a perfect region of the crystal. The thermal component of resistivity would then be

$$\rho_i(T) = \rho_i{}^\circ(T)(1 - \alpha), \quad \dots\dots\dots\dots\dots\dots (1)$$

where $\rho_i(T)$ is the resistance due to thermal vibrations, $\rho_i{}^\circ(T)$ is the value of $\rho_i(T)$ in the absence of the imperfections considered, and α is proportional to, and of the order of, the fraction of the volume of the crystal occupied by imperfections. On this basis we can understand the decrease of $\rho_i(T)$ with deformation.

II. PARTICULAR MODELS

In order to check whether the fractional volume α, as found for the observed deviations, is of reasonable order of magnitude it is necessary to assume some

* Division of Physics, C.S.I.R.O., University Grounds, Sydney.

model for the scattering imperfections. We shall consider two models : scattering by dislocations (disordered regions cylindrical) and scattering by stacking errors (disordered regions laminar).

Consider the following case, typical in regard to order of magnitude :

$$\delta\rho/\rho = 30 \text{ per cent.,} \quad \delta(d\rho/dT)/(d\rho/dT) = -2 \text{ per cent.,}$$

and ρ is about five times the resistance of copper, so that, from knowledge of the mean free path of electrons in copper, we find that the mean free path l due to scattering by the additional imperfections is 560 Å. The average distance between atoms is 3 Å.

Assuming that scattering is by dislocations, let there be N dislocation lines per cm.^{2} ; let r_1 be the radius of the cylindrical regions within which thermal displacements do not contribute to the scattering ; and let r_2 be the radius defining the scattering cross section, then

$$\pi N r_1^2 = \alpha = 0 \cdot 02,$$
$$2 N r_2 = 1/l = (5 \cdot 6 \times 10^{-6})^{-1}. \quad \dots \dots \dots \dots \dots \dots \quad (2)$$

If we now assume $r_1 = r_2 = r$, which is correct as regards order of magnitude,

$$r = \frac{2}{\pi} \times 0 \cdot 02 \times 5 \cdot 6 \times 10^{-6} = 7 \times 10^{-8} \text{ cm.,}$$
$$N = 1 \cdot 3 \times 10^{12} \text{ cm.}^{-2} \quad \dots \dots \dots \dots \dots \dots \dots \dots \dots \quad (3)$$

Both this effective radius and this dislocation density are of reasonable order of magnitude.

Alternatively, if it is assumed that scattering is by stacking errors of the type proposed by Broom (1952), similar agreement is obtained. Let the area of stacking error per unit volume be A, and let p be the probability of scattering for a Fermi electron passing through a stacking error. We assume that the effective volume of a stacking error is two atomic layers thick, i.e. 6 Å. Then

$$6 \times 10^{-8} A = \alpha = 0 \cdot 02,$$
$$Ap = (5 \cdot 6 \times 10^{-6})^{-1}. \quad \dots \dots \dots \dots \dots \dots \quad (4)$$

Hence $p = 0 \cdot 53$ and $A = 3 \times 10^{-5}$ cm.^{2}.

The magnitude of the scattering probability thus derived seems reasonable. An elementary wave-fitting calculation, assuming perpendicular incidence and a sudden phase change of $2\pi/3$ gives, after averaging over all phases, a reflection coefficient of $0 \cdot 5$. While the fortuitously close agreement with p is of no particular significance, the present picture is clearly self-consistent in regard to orders of magnitude.

The assumption that the volume occupied by imperfections does not contribute to the thermal scattering of electrons is therefore consistent with observations on the assumption of either of the two models.

III. References

BOAS, W., and NICHOLAS, J. F. (1953).—*Aust. J. Phys.* **6** : 116.
BROOM, T. (1952).—*Proc. Phys. Soc. Lond.* B **65** : 871.

"*The Photovoltaic Effect in Natural Lead Sulphide*"

By R. Lawrance

The graphical results of Figure 7 of this paper are referred to : If we plot log mobility v. log conductivity, we expect a straight line with a slope of 45°, assuming that the number of carriers remains constant over the conductivity range considered.   The slope of the straight line in Figure 7 is not 45°, and does not correctly describe the results presented in Table 3.
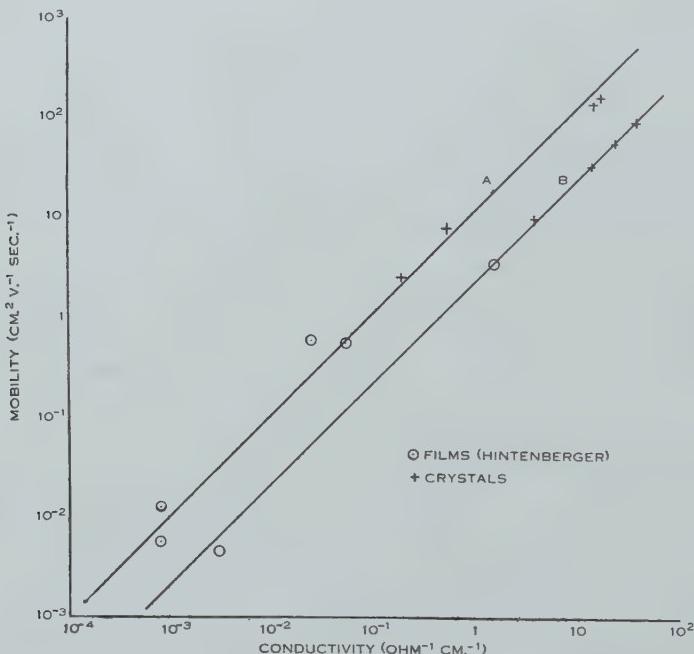
The figure reproduced here shows two groups of experimental data taken from Table 3 and from Hintenberger, for each of which the log mobility v. log conductivity is a straight line with a slope of 45°.   Between curve A and curve B there is a ratio of approximately 6 : 1 in the number of carriers.   In each group the number of carriers is constant within 10 per cent. for crystals.   The variation for films in curve A is about 50 per cent. and in curve B about 10 per cent. The groups cannot at present be identified by other characteristics of the samples, but their existence is suggested by the experimental results.   Obviously a systematic error in the measurements for the groups could account for their

existence. Furthermore, the number of samples investigated may not be regarded as large enough to establish this grouping. In any case, the grouping of the data in this way is consistent with the statement that the number of carriers is constant, at least in a less general sense than it was originally made.

The matter will be investigated further.

———

VOLUME 5, NUMBER 2, PAGES 266–87

" *The Distribution of the Discrete Sources of Cosmic Radio Radiation* "

By B. Y. Mills

The calculation performed on p. 277 is incorrect ; the first equation should read

$$p = P_0 n^{-4/3} \quad (\text{not } P_0 n^{-0.75}).$$

When the ensuing calculations are corrected it is found that the total power received in the galactic plane due to the integrated emission of the class I sources is reduced to approximately one-fifth. The conclusion in the paper that emission from the class I sources is inadequate to account for the observed general background radiation from the Galaxy is therefore strengthened. On p. 281, where the result of a similar calculation is given, no such error was made.

CONDUCTION OF HEAT IN THE SEMI–INFINITE SOLID, WITH A SHORT TABLE OF AN IMPORTANT INTEGRAL

By R. C. T. Smith*

Summary

The solution of a problem in heat conduction is expressed in terms of the integral $\int_0^U e^{-\alpha(1+u^2)} \dfrac{du}{1+u^2}$. This integral is tabulated for various values of α and U.

I. Introduction

The temperature distribution in a semi-infinite solid is determined for the following conditions : initial temperature zero, the plane $x=0$ held at the constant temperature θ for a time T then made impervious to heat.

II. Calculations and Results

To maintain the temperature θ at $x=0$ heat must be supplied at the rate $\theta K(\varkappa \pi t)^{-\frac{1}{2}}$ per unit time per unit area. The temperature distribution for $x>0$ will therefore be the same as in an infinite solid with a continuous plane source at $x=0$ liberating heat at the rate $2\theta K(\varkappa \pi t)^{-\frac{1}{2}}$ per unit time per unit area for $t<T$.

For $t<T$, the temperature is given by

$$v(x,\, t) = \frac{\theta}{\pi} \int_0^t e^{-x^2/4\varkappa(t-\tau)} \frac{d\tau}{\tau^{\frac{1}{2}}(t-\tau)^{\frac{1}{2}}},$$

that is,

$$v(x,\, t) = \theta \left\{ 1 - \mathrm{erf}\, \frac{x}{2\sqrt{\varkappa t}} \right\}, \quad \dots\dots\dots\dots\dots \quad (1)$$

while, for $t>T$,

$$v(x,\, t) = \frac{\theta}{\pi} \int_0^T e^{-x^2/4\varkappa(t-\tau)} \frac{d\tau}{\tau^{\frac{1}{2}}(t-\tau)^{\frac{1}{2}}}. \quad \dots\dots\dots \quad (2)$$

The temperature at $x=0$ is given by

$$v(0,\, t) = \begin{cases} \theta, & 0<t<T \\ \dfrac{2}{\pi}\theta \sin^{-1}\left(\dfrac{T}{t}\right)^{\frac{1}{2}}, & t>T \end{cases}. \quad \dots\dots \quad (3)$$

(For these results see Carslaw and Jaeger (1947, pp. 57, 222).)

The substitution

$$\tau = \frac{u^2}{1+u^2} t$$

* Department of Mathematics, New England University College, Armidale, N.S.W.

A

TABLE 1

$$\int_0^U e^{-\alpha(1+u^2)}\frac{du}{1+u^2}$$

α / U	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9	1·0
0·1	0·09015	0·08155	0·07376	0·06672	0·06035	0·05459	0·04938	0·04467	0·04040	0·03655
0·2	0·17838	0·16119	0·14566	0·13162	0·11894	0·10748	0·09713	0·08777	0·07931	0·07167
0·3	0·26295	0·23723	0·21403	0·19310	0·17422	0·15719	0·14182	0·12795	0·11545	0·10416
0·4	0·34254	0·30837	0·27761	0·24993	0·22501	0·20259	0·18240	0·16422	0·14786	0·13314
0·5	0·41626	0·37374	0·33557	0·30132	0·27058	0·24299	0·21822	0·19599	0·17603	0·15812
0·6	0·48366	0·43290	0·38751	0·34692	0·31062	0·27814	0·24908	0·22308	0·19982	0·17900
0·7	0·54464	0·48580	0·43340	0·38673	0·34515	0·30809	0·27507	0·24562	0·21937	0·19596
0·8	0·59940	0·53264	0·47347	0·42100	0·37447	0·33317	0·29652	0·26398	0·23508	0·20940
0·9	0·64829	0·57380	0·50812	0·45017	0·39903	0·35385	0·31393	0·27864	0·24742	0·21979
1·0	0·69176	0·60975	0·53784	0·47475	0·41935	0·37066	0·32783	0·29013	0·25693	0·22765
1·1	0·73033	0·64100	0·56318	0·49529	0·43600	0·38415	0·33877	0·29900	0·26411	0·23348
1·2	0·76448	0·66808	0·58465	0·51232	0·44950	0·39486	0·34726	0·30574	0·26946	0·23772
1·3	0·79470	0·69148	0·60276	0·52634	0·46035	0·40327	0·35377	0·31078	0·27336	0·24074
1·4	0·82144	0·71164	0·61797	0·53781	0·46901	0·40979	0·35870	0·31449	0·27616	0·24286
1·5	0·84509	0·72899	0·63069	0·54714	0·47586	0·41482	0·36238	0·31720	0·27815	0·24431
1·6	0·86601	0·74388	0·64130	0·55469	0·48123	0·41865	0·36511	0·31914	0·27953	0·24530
1·7	0·88454	0·75666	0·65010	0·56076	0·48542	0·42153	0·36710	0·32051	0·28048	0·24595
1·8	0·90095	0·76759	0·65739	0·56562	0·48865	0·42369	0·36854	0·32147	0·28112	0·24638
1·9	0·91549	0·77693	0·66340	0·56948	0·49114	0·42529	0·36956	0·32213	0·28154	0·24665
2·0	0·92838	0·78491	0·66833	0·57254	0·49303	0·42646	0·37029	0·32258	0·28182	0·24682
2·5	0·97404	0·81009	0·68228	0·58029	0·49735	0·42887	0·37165	0·32335	0·28225	0·24707
3·0	0·99920	0·82094	0·68698	0·58234	0·49825	0·42928	0·37183	0·32343	0·28229	0·24708
∞	1·02843	0·82795	0·68892	0·58291	0·49843	0·42933	0·37184	0·32343	0·28229	0·24709

α / U	1·1	1·2	1·3	1·4	1·5	1·6	1·7	1·8	1·9	2·0
0·1	0·03306	0·02990	0·02705	0·02446	0·02213	0·02002	0·01811	0·01638	0·01481	0·01340
0·2	0·06477	0·05853	0·05289	0·04779	0·04319	0·03903	0·03527	0·03187	0·02880	0·02603
0·3	0·09398	0·08479	0·07651	0·06903	0·06228	0·05620	0·05071	0·04575	0·04128	0·03725
0·4	0·11988	0·10794	0·09720	0·08752	0·07881	0·07097	0·06391	0·05756	0·05183	0·04668
0·5	0·14203	0·12759	0·11462	0·10297	0·09251	0·08312	0·07468	0·06711	0·06030	0·05419
0·6	0·16036	0·14368	0·12875	0·11538	0·10340	0·09268	0·08308	0·07448	0·06678	0·05988
0·7	0·17508	0·15645	0·13982	0·12499	0·11174	0·09992	0·08936	0·07993	0·07150	0·06398
0·8	0·18657	0·16628	0·14824	0·13219	0·11790	0·10519	0·09387	0·08379	0·07481	0·06680
0·9	0·19532	0·17365	0·15444	0·13741	0·12230	0·10889	0·09699	0·08642	0·07702	0·06867
1·0	0·20183	0·17903	0·15889	0·14109	0·12535	0·11141	0·09907	0·08814	0·07844	0·06985
1·1	0·20655	0·18286	0·16200	0·14361	0·12739	0·11307	0·10041	0·08923	0·07933	0·07056
1·2	0·20991	0·18553	0·16411	0·14529	0·12872	0·11412	0·10125	0·08989	0·07985	0·07098
1·3	0·21225	0·18734	0·16552	0·14638	0·12956	0·11478	0·10176	0·09028	0·08016	0·07122
1·4	0·21385	0·18855	0·16643	0·14706	0·13008	0·11517	0·10205	0·09051	0·08033	0·07134
1·5	0·21492	0·18933	0·16700	0·14749	0·13039	0·11540	0·10222	0·09063	0·08042	0·07141
1·6	0·21562	0·18983	0·16736	0·14774	0·13057	0·11552	0·10231	0·09070	0·08046	0·07144
1·7	0·21607	0·19014	0·16757	0·14789	0·13067	0·11559	0·10236	0·09073	0·08049	0·07146
1·8	0·21636	0·19033	0·16770	0·14797	0·13073	0·11563	0·10239	0·09075	0·08050	0·07147
1·9	0·21653	0·19045	0·16777	0·14802	0·13076	0·11565	0·10240	0·09075	0·08050	0·07147
2·0	0·21664	0·19051	0·16781	0·14804	0·13078	0·11566	0·10240	0·09076	0·08051	0·07147
2·5	0·21678	0·19059	0·16786	0·14807	0·13079	0·11567	0·10241	0·09076	0·08051	0·07147
3·0	0·21679	0·19059	0·16786	0·14807	0·13079	0·11567	0·10241	0·09076	0·08051	0·07147
∞	0·21679	0·19059	0·16786	0·14807	0·13079	0·11567	0·10241	0·09076	0·08051	0·07147

TABLE 1 (*Continued*)

α U	2·5	3·0	4·0	5·0	α U	2·5	3·0	4·0	5·0
0·1	0·00811	0·00491	0·00180	0·00066	1·1	0·03956	0·02240	0·00734	0·00246
0·2	0·01568	0·00945	0·00343	0·00125	1·2	0·03969	0·02244	0·00735	,,
0·3	0·02228	0·01333	0·00477	0·00171	1·3	0·03976	0·02246	,,	,,
0·4	0·02766	0·01640	0·00577	0·00204	1·4	0·03979	0·02247	,,	,,
0·5	0·03178	0·01866	0·00645	0·00224	1·5	0·03980	,,	,,	,,
0·6	0·03475	0·02021	0·00688	0·00236	1·6	0·03981	,,	,,	,,
0·7	0·03677	0·02120	0·00712	0·00242	1·7	0·03981	,,	,,	,,
0·8	0·03806	0·02180	0·00724	0·00244	1·8	0·03982	,,	,,	,,
0·9	0·03885	0·02213	0·00730	0·00245	1·9	,,	,,	,,	,,
1·0	0·03931	0·02231	0·00733	0·00246	2·0	,,	,,	,,	,,
					∞	,,	,,	,,	,,

reduces the integral (**2**) to the form

$$v(x,\,t)=\frac{2\theta}{\pi}\int_0^{\sqrt{T/(t-T)}}e^{-x^2(1+u^2)/4\varkappa t}\,\frac{du}{1+u^2}, \quad \ldots\ldots \text{ (4)}$$

that is, to the form

$$v(x,\,t)=\frac{2\theta}{\pi}\int_0^{U}e^{-\alpha(1+u^2)}\,\frac{du}{1+u^2}, \quad \ldots\ldots\ldots\ldots \text{ (5)}$$

with

$$U=\sqrt{\frac{T}{t-T}}, \qquad \alpha=\frac{x^2}{4\varkappa t}. \quad \ldots\ldots\ldots\ldots \text{ (6)}$$

The integral (**5**) has been calculated from rearranged forms of the series expansions quoted by Lightfoot (1930)

$$\int_0^{U}e^{-\alpha(1+u^2)}\frac{du}{1+u^2}=e^{-\alpha}\sum_{n=0}^{\infty}\frac{(-1)^n U^{2n+1}}{2n+1}\left\{1+\frac{\alpha}{1!}+\cdots+\frac{\alpha^n}{n!}\right\}$$

for $U\leqslant1$,

$$\int_0^{U}e^{-\alpha(1+u^2)}\frac{du}{1+u^2}=\frac{\pi}{2}\,\{1-\mathrm{erf}\,\sqrt{\alpha}\,\mathrm{erf}\,U\sqrt{\alpha}\}$$

$$-e^{-\alpha U^2}\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n+1)\,U^{2n+1}}\left\{1+\frac{\alpha U^2}{1!}+\cdots+\frac{(\alpha U^2)^n}{n!}\right\}$$

for $U>1$,

and the results are given in Table 1. The calculations were made using six figures and differencing did not show errors greater than two in the last figure. The results were then rounded off to five figures so that the final table should not have rounding-off errors greater than $0\cdot7$ in the last figure.

Since little arithmetic is involved in passing from the integral to the temperature, no table of temperature distribution is given. The results are presented graphically in Figure 1, $x/(2\sqrt{\varkappa t})$ being plotted against $\sqrt{t/T}$ for fixed values of $v(x, t)/\theta$.
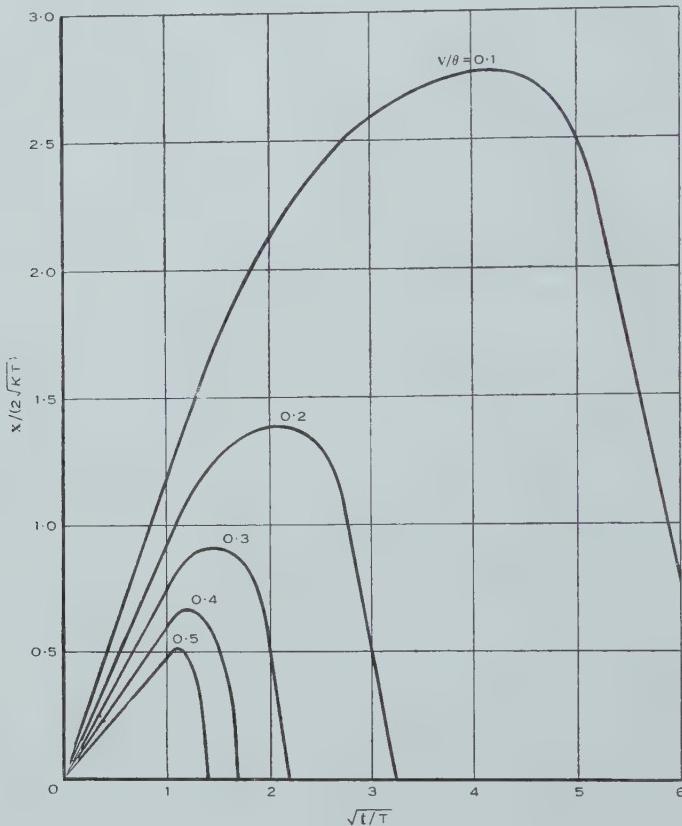

Fig. 1

III. References

CARSLAW, H. S., and JAEGER, J. C. (1947).—" Conduction of Heat in Solids." (Oxford Univ. Press.)

LIGHTFOOT, N. M. H. (1930).—The solidification of molten steel. *Proc. Lond. Math. Soc.* (2) **31** : 97-116.

ON THE STANDARD ERRORS IN THE FITTING OF POLYNOMIALS TO UNEQUALLY SPACED OBSERVATIONS

By P. G. Guest*

Summary

From the observed values of the independent variable two parameters are derived which specify the departure from uniform spacing. Expressions are obtained for the standard errors of the coefficients and fitted values in terms of these parameters, and numerical tables for the estimation of the errors are given. It is shown that the errors calculated in this way lie within a few per cent. of the exact least squares values for polynomials of the first and second degree, but when the polynomial is of the third degree the deviations may be much greater.

I. Introduction

The problem of fitting polynomials to equally spaced observations has been very thoroughly treated by a number of authors. When the observations are not equally spaced the problem is much more difficult. In fact, each set of observations requires its own separate treatment, and no general information concerning the values of the orthogonal polynomials $T_j(x)$ is available. Information of this type is essential if an adequate discussion is to be given of approximate methods of curve fitting such as those described in an earlier paper (Guest 1952). All that could be done in that paper was to assume that the values $\Sigma T_j{}^2(x)$ did not differ greatly from those in the equally spaced case. The efficiencies of the methods of grouping described there are calculated on this assumption, and hence may be in error if the spacing is markedly non-uniform.

The aim of the present paper is to prepare the way for a new attack on this problem by discussing the standard errors of the least squares polynomial coefficients and fitted values when the observations are unequally spaced. The procedure adopted is to characterize any particular set of observations by two parameters $\varkappa_2$, $\varkappa_3$. The parameter $\varkappa_2$ is a measure of the departure of the independent variable x from symmetry about the central value, while the parameter $\varkappa_3$ is a measure of the relative concentration of the observations towards the central values of x as opposed to the extreme values. It is shown that $\Sigma T_j{}^2(x)$ can be expressed approximately as a function of $\varkappa_2$, $\varkappa_3$, and tables have been prepared from which these functions may be obtained. Tables for the calculation of the standard errors of the fitted values are also given.

Although the quantities tabulated were calculated for use in future theoretical discussions, they may also be of use in practical examples, either for the rough calculation of the standard errors or as a check on the values

* Physics Department, University of Sydney.

obtained by the usual methods. Some examples showing how the values obtained by the present method agree with the exact least squares values will be given.

Unfortunately, the treatment in terms of the parameters $\varkappa_2$, $\varkappa_3$ is not adequate for all possible sets of data. In certain cases additional higher order parameters $\varkappa_4$, $\varkappa_5$ are required. The higher the degree of the polynomial to be fitted the more important will these additional parameters be. However, it is found that the treatment given here is adequate for practically all cases in which the curve is of the first or second degree, and for a large proportion of the cases in which the curve is of the third degree.

II. The Smoothing of the Points of Observation

If the values of the independent variable x at the n points of observation are arranged in order of magnitude, each observation may be identified by a number ε giving its position in the sequence, ε taking the integral or half-integral values from $+\frac{1}{2}(n-1)$ to $-\frac{1}{2}(n-1)$. A point of observation can then be represented by the symbol $x(\varepsilon)$. In the present discussion the system of points $x(\varepsilon)$ will be replaced by a smoothed-out system $X(\varepsilon)$ obtained by fitting a curve of the third degree in ε to the values $x(\varepsilon)$. The smoothed-out system of points is given by the equation

$$X(\varepsilon) = k_0 + k_1 T_1(\varepsilon) + k_2 T_2(\varepsilon) + k_3 T_3(\varepsilon), \quad \ldots\ldots\ldots\ldots \text{ (1)}$$

where

$$k_j = \sum_\varepsilon T_j(\varepsilon) x(\varepsilon) / \sum_\varepsilon T_j{}^2(\varepsilon), \quad \ldots\ldots\ldots\ldots\ldots\ldots \text{ (2)}$$

and $T_j(\varepsilon)$ is the orthogonal polynomial of degree j in ε with leading coefficient unity.

The variable $X(\varepsilon)$ can be transformed by a change of origin and scale to give a new variable

$$\xi(\varepsilon) = \varphi\{X(\varepsilon) - k_0\}. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{ (3)}$$

Then equation (1) becomes

$$\xi(\varepsilon) = \varkappa_1 \tau_1(\varepsilon) + \varkappa_2 \tau_2(\varepsilon) + 2\varkappa_3 \tau_3(\varepsilon), \quad \ldots\ldots\ldots\ldots\ldots \text{ (4)}$$

where

$$\tau_j(\varepsilon) = n^{-j+1} T_j(\varepsilon), \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{ (5)}$$

and

$$\varkappa_1 = \varphi k_1, \quad \varkappa_2 = n\varphi k_2, \quad \varkappa_3 = \tfrac{1}{2}n^2\varphi k_3. \quad \ldots\ldots\ldots\ldots \text{ (6)}$$

The advantage of this notation is that the coefficients $\varkappa_j$ are all of the same order of magnitude.

Since it is desired to compare the smoothed set $\xi(\varepsilon)$ with the equally spaced set ε, the scale factor φ in (3) should be chosen so that the mean interval between successive observations is unity. The mean interval is

$$\frac{1}{n-1}\left\{\xi\left(+\frac{n-1}{2}\right) - \xi\left(-\frac{n-1}{2}\right)\right\} = \varphi\left\{k_1 + \frac{(n-2)(n-3)}{10}k_3\right\},$$

and this would give for φ the value $\left\{k_1 + \frac{1}{10}(n-2)(n-3)k_3\right\}^{-1}$. However, the

term $(n-2)(n-3)$ is found to be inconvenient in subsequent calculations, and the simpler form

$$\varphi = \left\{ k_1 + \frac{1}{10} n^2 k_3 \right\}^{-1} \quad \dots\dots\dots\dots\dots \quad (7)$$

will be adopted. From equations (6) and (7) it follows that

$$\varkappa_1 = 1 - \frac{1}{5} \varkappa_3. \quad \dots\dots\dots\dots\dots\dots\dots \quad (8)$$

The original set of points $x(\varepsilon)$ is thus replaced by a smoothed-out set characterized by three parameters φ, $\varkappa_2$, $\varkappa_3$. The parameter φ gives the scale factor, while the parameters $\varkappa_2$, $\varkappa_3$ specify the departure from uniform spacing.

III. The Parameters $\varkappa_2$, $\varkappa_3$

Values of k_1, k_2, k_3 may be found by one of the conventional least squares methods for equally spaced data, x being treated as the dependent variable and ε as the independent variable.

A more rapid procedure is to use the method of weighted grouping (Guest 1951). k_2, k_3 are identical with the quantities denoted by a_2, a_3 in the reference quoted, while

$$k_1 - a_1 + \frac{1}{40} \{n^2 - 5(\nu + r)^2\} a_3.$$

when $n = 3\nu - r$, ν being an integer and r having the value -1, 0, or $+1$.

Very often comparatively rough approximations will suffice. If X_1, X_2, X_3, X_4, X_5 denote the values of X for which ε takes the values $+\frac{1}{2}(n-1)$, $+\frac{1}{4}(n-1)$, 0, $-\frac{1}{4}(n-1)$, $-\frac{1}{2}(n-1)$, then from equation (1)

$$\frac{X_1 + X_5 - 2X_3}{(n-1)^2} = \frac{1}{2} k_2, \quad \dots\dots\dots\dots\dots \quad (9)$$

$$\frac{X_1 - X_5 - 2(X_2 - X_4)}{(n-1)^3} = \frac{3}{16} k_3. \quad \dots\dots\dots\dots \quad (10)$$

Also

$$\frac{X_1 - X_5}{n-1} = k_1 + \frac{1}{10}(n-2)(n-3)k_3$$

$$= \varphi^{-1} - \frac{1}{10} k_3(5n-6).$$

Hence, if x_i denote the observed values x corresponding to the same five values of ε, the values of the parameters can be estimated from the following formulae :

$$\left. \begin{array}{l} k_2 = \dfrac{2(x_1 + x_5 - 2x_3)}{(n-1)^2}, \\[2ex] k_3 = \dfrac{16\{(x_1 - x_5) - 2(x_2 - x_4)\}}{3(n-1)^3}, \end{array} \right\} \quad \dots\dots\dots \quad (11)$$

$$\varphi^{-1} = \frac{x_1 - x_5}{n-1} + \frac{1}{2} n k_3. \quad \dots\dots\dots\dots\dots \quad (12)$$

$$\left. \begin{array}{l} \varkappa_2 = n\varphi k_2, \\[1ex] \varkappa_3 = \frac{1}{2} n^2 \varphi k_3. \end{array} \right\} \quad \dots\dots\dots\dots\dots \quad (13)$$

The significance of the parameters x_2, x_3 can be brought out by combining equations (11), (12), and (13) to give the following approximate equations :

$$\frac{x_1 - x_3}{x_1 - x_5} \doteqdot \frac{2 + x_2}{4}, \quad \dots\dots\dots\dots\dots \quad (14)$$

$$\frac{x_2 - x_4}{x_1 - x_5} \doteqdot \frac{8 - 3x_3}{16}. \quad \dots\dots\dots\dots\dots \quad (15)$$

Thus x_2 is a measure of the departure from symmetry about the central value x_3 ($\varepsilon = 0$). x_3 is a measure of the relative concentration of the observations towards the centre of the range. For the equally spaced case $x_2 = x_3 = 0$. When $x_2 = +1$, the first half of the observations (for which ε is positive) is spread over three-quarters of the range of x. When $x_3 = +4/3$, the central half of the observations (for which $|\varepsilon|$ is less than $\frac{1}{4}(n-1)$) is confined to a quarter of the range of x.

There does not appear to be any very simple criterion which fixes the ranges of values of x_2, x_3 likely to be encountered in practical examples. However, one condition which suggests itself is that the difference $\Delta\xi(\varepsilon)$ should always be positive, that is, the sequence of the smoothed values $X(\varepsilon)$ should be the same as the sequence of the actual values $x(\varepsilon)$.

The finite difference $\Delta\xi(\varepsilon)$ can be obtained from the approximate formula

$$\Delta\xi(\varepsilon) \doteqdot x_1 + 2n^{-1}x_2\varepsilon + 2n^{-2}x_3(3\varepsilon^2 - 3n^2/20),$$

or

$$\Delta\xi \doteqdot 1 + x_2 e + \tfrac{1}{2}x_3(3e^2 - 1), \quad \dots\dots\dots\dots\dots \quad (16)$$

where $e = 2\varepsilon/n$. It is required to find the condition that $\Delta\xi$ should never be negative in the range $-1 < e < +1$.

If x_3 is positive there is a minimum value of $\Delta\xi$ at $e = -x_2/3x_3$. When $|x_2| < 3x_3$ this minimum occurs in the range $-1 < e < +1$, and has the value

$$1 - \frac{1}{6}\frac{x_2^2}{x_3} - \tfrac{1}{2}x_3.$$

Hence, if this is to be positive,

$$x_2^2 < 3x_3(2 - x_3). \quad \dots\dots\dots\dots\dots \quad (17)$$

When $|x_2| > 3x_3$, or when x_3 is negative, the least value of $\Delta\xi$ in the range $-1 < e < +1$ will occur at one end of the range, the value then being

$$1 - |x_2| + x_3.$$

If this is to be positive,

$$|x_2| < 1 + x_3. \quad \dots\dots\dots\dots\dots \quad (18)$$

Conditions (17) and (18) may be summarized conveniently as follows :

$$\left.\begin{array}{ll} \text{for } -1 \ \leqslant x_3 \leqslant +0\cdot5, & |x_2| \leqslant 1 + x_3 \,; \\ \text{for } +0\cdot5 \leqslant x_3 \leqslant +2, & |x_2| \leqslant \{3x_3(2 - x_3)\}^{\frac{1}{2}}. \end{array}\right\} \ \dots (19)$$

If the values of x_2, x_3 lie outside these ranges, then it would certainly be necessary to include terms of higher degree in equation (1).

IV. The Orthogonal Polynomials $T_j(\xi)$

The orthogonal polynomials $T_j(\xi)$, together with the associated coefficients α_{jk} and β_{jk} defined by the equations

$$T_j(\xi) = \xi^j + \sum_0^{j-1} \alpha_{jk} T_k(\xi), \quad \dots\dots\dots\dots \quad (20)$$

$$T_j(\xi) = \sum_0^j \beta_{jk} \xi^k, \quad \dots\dots\dots\dots\dots \quad (21)$$

are used in the calculation of the standard errors. Hence it is necessary to obtain formulae for which these quantities may be calculated in terms of the parameters $\varkappa_2$, $\varkappa_3$.

A convenient method of procedure is to expand ξ^j in terms of the orthogonal polynomials $\tau_j(\varepsilon)$, since this permits the rapid calculation of the sums $\Sigma T_j{}^2(\xi)$. Accordingly ξ^j is written in the form

$$n^{-(j-1)}\xi^j = \sum_{k=1}^{3j} \varkappa_{jk}\tau_k(\varepsilon) + \sum_{m=0}^{j-2} n^{-m}\varkappa_{j-m,0}\xi^m. \quad \dots\dots \quad (22)$$

Now the left-hand side of this equation can be rewritten as

$$n^{-1}\xi\{n^{-(j-2)}\xi^{j-1}\} = n^{-1}\sum_1^{3(j-1)} \varkappa_{j-1,k}\tau_k(\varepsilon)\sum_1^3 \varkappa_{1m}\tau_m(\varepsilon) + \sum_0^{j-3} n^{-(m-1)}\varkappa_{j-m,0}\xi^{m+1},$$

and so

$$\sum_0^{3j} \varkappa_{jk}\tau_k(\varepsilon) = n^{-1}\sum_1^{3(j-1)} \varkappa_{j-1,k}\tau_k(\varepsilon)\sum_1^3 \varkappa_{1m}\tau_m(\varepsilon). \quad \dots\dots\dots \quad (23)$$

The expression on the right may be evaluated by using the recurrence relation

$$n^{-1}\varepsilon\tau_j(\varepsilon) = \tau_{j+1}(\varepsilon) + \rho_j\tau_{j-1}(\varepsilon), \quad \dots\dots\dots\dots \quad (24)$$

where

$$\rho_j = \frac{j^2(1-j^2/n^2)}{4(4j^2-1)}.$$

To simplify the calculations the term j^2/n^2 will be neglected, and the form

$$\rho_j = \frac{j^2}{4(4j^2-1)} \quad \dots\dots\dots\dots\dots \quad (25)$$

will be used.

The expressions required for the calculation of $\varkappa_{jk}$ for values of j from 1 to 3 are listed in Table 1. The explicit expressions for these coefficients in terms of $\varkappa_2$, $\varkappa_3$ are given in Table 2.

From equations (20) and (22), $\tau_j(\xi)$ can be expressed in the form

$$\tau_j(\xi) = \sum_1^{3j} \varkappa_{jk}\tau_k(\varepsilon) + \sum_1^{j-1} \mu_{jm}\tau_m(\xi). \quad \dots\dots\dots \quad (26)$$

Also the sum $\Sigma\tau_j{}^2(\xi)$ may be written as

$$\Sigma\tau_j{}^2(\xi) = \mu_{jj}\Sigma\tau_j{}^2(\varepsilon). \quad \dots\dots\dots\dots \quad (27)$$

It remains to calculate μ_{jm}, μ_{jj}.

If equation (24) is multiplied by $\tau_{j-1}(\varepsilon)$, it follows that

$$\rho_j \sum_\varepsilon \tau_{j-1}{}^2(\varepsilon) = \sum_\varepsilon \tau_j(\varepsilon)\{n^{-1}\varepsilon\tau_{j-1}(\varepsilon)\}$$

$$-\sum_\varepsilon \tau_j{}^2(\varepsilon).$$

TABLE 1

QUANTITIES REQUIRED IN THE CALCULATION OF THE COEFFICIENTS $\varkappa_{jk}$

ρ_j	$\rho_j = \dfrac{j^2}{4(4j^2-1)}$	$R_j = \prod_1^j \rho_k = \dfrac{(j!)^4}{(2j)!(2j+1)!}$
$\tau_j(\varepsilon)\tau_k(\varepsilon)$	$n^{-1}\tau_1(\varepsilon)\tau_j(\varepsilon) = \tau_{j+1}(\varepsilon) + \rho_j\tau_{j-1}(\varepsilon)$ $n^{-1}\tau_2(\varepsilon)\tau_j(\varepsilon) = \tau_{j+2}(\varepsilon) + \left(\rho_{j+1} + \rho_j - \dfrac{1}{12}\right)\tau_j(\varepsilon) + \rho_j\rho_{j-1}\tau_{j-2}(\varepsilon)$ $n^{-1}\tau_3(\varepsilon)\tau_j(\varepsilon) = \tau_{j+3}(\varepsilon) + \left(\rho_{j+2} + \rho_{j+1} + \rho_j - \dfrac{3}{20}\right)\tau_{j+1}(\varepsilon)$ $+\rho_j\left(\rho_{j+1} + \rho_j + \rho_{j-1} - \dfrac{3}{20}\right)\tau_{j-1}(\varepsilon) + \rho_j\rho_{j-1}\rho_{j-2}\tau_{j-3}(\varepsilon)$	

TABLE 2

THE COEFFICIENTS $\varkappa_{jk}$

$\varkappa_{11} = 1 - \dfrac{1}{5}\varkappa_3$	$\varkappa_{30} = n^{-3}\Sigma\xi^3$
$\varkappa_{12} = \varkappa_2$	$\varkappa_{31} = \dfrac{1}{15} - \dfrac{1}{70}\varkappa_3 + \dfrac{1}{350}\varkappa_3{}^2 - \dfrac{2}{5775}\varkappa_3{}^3 + \varkappa_2{}^2\left(\dfrac{13}{630} - \dfrac{2}{1575}\varkappa_3\right)$
$\varkappa_{13} = 2\varkappa_3$	$\varkappa_{32} = \varkappa_2\left(\dfrac{13}{42} - \dfrac{4}{105}\varkappa_3 + \dfrac{2}{1155}\varkappa_3{}^2 + \dfrac{2}{315}\varkappa_2{}^2\right)$
$\varkappa_{20} = n^{-2}\Sigma\xi^2$	$\varkappa_{33} = 1 - \dfrac{3}{55}\varkappa_3{}^2 + \dfrac{34}{5005}\varkappa_3{}^3 + \varkappa_2{}^2\left(\dfrac{1}{3} - \dfrac{8}{495}\varkappa_3\right)$
$\varkappa_{21} = \dfrac{2}{15}\varkappa_2\left(1 - \dfrac{1}{14}\varkappa_3\right)$	$\varkappa_{34} = \varkappa_2\left(3 + \dfrac{1}{11}\varkappa_3 - \dfrac{5}{143}\varkappa_3{}^2 + \dfrac{1}{11}\varkappa_2{}^2\right)$
$\varkappa_{22} = 1 - \dfrac{1}{7}\varkappa_3 + \dfrac{1}{21}\varkappa_2{}^2$	$\varkappa_{35} = 6\varkappa_3 - \dfrac{15}{13}\varkappa_3{}^2 + \dfrac{3}{65}\varkappa_3{}^3 + \varkappa_2{}^2\left(3 - \dfrac{1}{13}\varkappa_3\right)$
$\varkappa_{23} = 2\varkappa_2\left(1 - \dfrac{1}{9}\varkappa_3\right)$	$\varkappa_{36} = \varkappa_2\left(12\varkappa_3 - \dfrac{7}{5}\varkappa_3{}^2 + \varkappa_2{}^2\right)$
$\varkappa_{24} = 4\varkappa_3 - \dfrac{7}{11}\varkappa_3{}^2 + \varkappa_2{}^2$	$\varkappa_{37} = 12\varkappa_3{}^2 - \dfrac{30}{17}\varkappa_3{}^3 + 6\varkappa_2{}^2\varkappa_3$
$\varkappa_{25} = 4\varkappa_2\varkappa_3$	$\varkappa_{38} = 12\varkappa_2\varkappa_3{}^2$
$\varkappa_{26} = 4\varkappa_3{}^2$	$\varkappa_{39} = 8\varkappa_3{}^3$

and so

where

$$\sum\tau_j{}^2(\varepsilon) = (R_j/R_m)\Sigma\tau_m{}^2(\varepsilon), \quad \ldots\ldots\ldots\ldots (28)$$

$$R_j = \prod_1^j \rho_k. \quad \ldots\ldots\ldots\ldots\ldots\ldots (29)$$

To calculate μ_{jm}, equation (26) is multiplied by $\tau_m(\xi)$ and summed over ξ, giving

$$\mu_{jm}\Sigma\tau_m^2(\xi) = -\Sigma_\xi\{\overset{3j}{\underset{1}{\Sigma}}\varkappa_{jk}\tau_k(\varepsilon)\tau_m(\xi)\} \quad\dots\dots\dots\dots\dots\dots\dots\dots \quad (30)$$

$$= -\Sigma_\xi\{\overset{3j}{\underset{1}{\Sigma}}\varkappa_{jk}\tau_k(\varepsilon)\}\{\overset{3m}{\underset{1}{\Sigma}}\varkappa_{mq}\tau_q(\varepsilon) + \overset{m-1}{\underset{1}{\Sigma}}\mu_{mr}\tau_r(\xi)\}$$

$$= -\{\overset{3m}{\underset{1}{\Sigma}}\varkappa_{jk}\varkappa_{mk}\Sigma\tau_k^2(\varepsilon) - \overset{m-1}{\underset{1}{\Sigma}}\mu_{mr}\mu_{jr}\Sigma\tau_r^2(\xi)\},$$

and so

$$\mu_{jm} = -\{\overset{3m}{\underset{1}{\Sigma}}\varkappa_{jk}\varkappa_{mk}R_k - \overset{m-1}{\underset{1}{\Sigma}}\mu_{jk}\mu_{mk}\mu_{kk}R_k\}/\mu_{mm}R_m. \quad\dots\dots (31)$$

To calculate μ_{jj}, equation (26) is squared, and, from (30), it follows that

$$\Sigma\tau_j^2(\xi) = \overset{3j}{\underset{1}{\Sigma}}\varkappa_{jk}^2\Sigma\tau_k^2(\varepsilon) - \overset{j-1}{\underset{1}{\Sigma}}\mu_{jm}^2\Sigma\tau_m^2(\xi).$$

Hence

$$\mu_{jj} = \{\overset{3j}{\underset{1}{\Sigma}}\varkappa_{jk}^2R_k - \overset{j-1}{\underset{1}{\Sigma}}\mu_{jm}^2\mu_{mm}R_m\}/R_j. \quad\dots\dots\dots\dots (32)$$

In Table 3 are listed the sums $\underset{m}{\Sigma}\varkappa_{jm}\varkappa_{km}R_m$ as explicit functions of $\varkappa_2$, $\varkappa_3$, and also the formulae for deriving μ_{jm}, μ_{jj} from these sums, for values of j, k from 1 to 3. In Table 4 the numerical values μ_{jj}, μ_{jm} are tabulated for the range $\varkappa_2^2 = 0(0\cdot25)1\cdot0(0\cdot5)2\cdot0$, $\varkappa_3 = -1\cdot0(0\cdot2)+2\cdot0$.

<div align="center">

TABLE 3

FORMULAE FOR THE DERIVATION OF THE QUANTITIES μ_{jj}, μ_{jm}

</div>

$$\overset{3}{\underset{1}{\Sigma}}\varkappa_{1j}^2R_j = [1 - 0\cdot400000\varkappa_3 + 0\cdot057143\varkappa_3^2 + 0\cdot066667\varkappa_2^2]R_1$$

$$\overset{3}{\underset{1}{\Sigma}}\varkappa_{1j}\varkappa_{2j}R_j = 0\cdot200000\varkappa_2[1 - 0\cdot142857\varkappa_3 + 0\cdot015873\varkappa_2^2]R_1$$

$$\left\{\begin{array}{l}\overset{6}{\underset{1}{\Sigma}}\varkappa_{2j}^2R_j \quad [(1 - 0\cdot285714\varkappa_3 + 0\cdot085714\varkappa_3^2 - 0\cdot020779\varkappa_3^3 + 0\cdot001912\varkappa_3^4) \\[4pt] \qquad\qquad + \varkappa_2^2(0\cdot619048 - 0\cdot076190\varkappa_3 + 0\cdot003463\varkappa_3^2) + \varkappa_2^4(0\cdot006349)]R_2 \\[8pt] \overset{3}{\underset{1}{\Sigma}}\varkappa_{1j}\varkappa_{3j}R_j\end{array}\right.$$

$$\overset{6}{\underset{1}{\Sigma}}\varkappa_{2j}\varkappa_{3j}R_j \quad 0\cdot571429\varkappa_2[(1 - 0\cdot150000\varkappa_3 + 0\cdot013636\varkappa_3^2 - 0\cdot001598\varkappa_3^3)$$
$$\qquad\qquad + \varkappa_2^2(0\cdot205556 - 0\cdot017677\varkappa_3 + 0\cdot000178\varkappa_3^2) + \varkappa_2^4(0\cdot001178)]R_2$$

$$\overset{9}{\underset{1}{\Sigma}}\varkappa_{3j}^2R_j = [(2\cdot037037 - 0\cdot444444\varkappa_3 + 0\cdot171717\varkappa_3^2 - 0\cdot071736\varkappa_3^3 + 0\cdot017027\varkappa_3^4$$
$$\qquad\qquad - 0\cdot002301\varkappa_3^5 + 0\cdot000136\varkappa_3^6)$$
$$\qquad\qquad + \varkappa_2^2(3\cdot370370 - 0\cdot397306\varkappa_3 + 0\cdot030947\varkappa_3^2 - 0\cdot005565\varkappa_3^3 + 0\cdot000383\varkappa_3^4)$$
$$\qquad\qquad + \varkappa_2^4(0\cdot342312 - 0\cdot025273\varkappa_3 + 0\cdot000464\varkappa_3^2) + \varkappa_2^6(0\cdot001404)]R_3$$

$$\mu_{11} = \Sigma\varkappa_{1j}^2R_j/R_1$$
$$\mu_{21} = -\{\Sigma\varkappa_{1j}\varkappa_{2j}R_j/R_1\}/\mu_{11}$$
$$\mu_{22} = \{\Sigma\varkappa_{2j}^2R_j/R_2\} - 15\mu_{21}^2\mu_{11}$$
$$\mu_{31} = -\{\Sigma\varkappa_{1j}\varkappa_{3j}R_j/R_2\}/15\mu_{11}$$
$$\mu_{32} = -\{(\Sigma\varkappa_{2j}\varkappa_{3j}R_j/R_2) - 15\mu_{31}\mu_{21}\mu_{11}\}/\mu_{22}$$
$$\mu_{33} = \{\Sigma\varkappa_{3j}^2R_j/R_3\} - (140/9)\mu_{32}^2\mu_{22} - (700/3)\mu_{31}^2\mu_{11}$$

P. G. GUEST

TABLE 4

NUMERICAL VALUES OF μ_{jj}, μ_{jm}

μ_{11}

x_3 \ $x_2{}^2$	0	0·25	0·50	0·75	1·00	1·50	2·00
−1·0	1·457	1·474	1·490	1·507	1·524	1·557	1·590
−0·8	1·357	1·373	1·390	1·407	1·423	1·457	1·490
−0·6	1·261	1·277	1·294	1·311	1·327	1·361	1·394
−0·4	1·169	1·186	1·202	1·219	1·236	1·269	1·302
−0·2	1·082	1·099	1·116	1·132	1·149	1·182	1·216
0	1·000	1·017	1·033	1·050	1·067	1·100	1·133
+0·2	0·922	0·939	0·956	0·972	0·989	1·022	1·056
+0·4	0·849	0·866	0·882	0·899	0·916	0·949	0·982
+0·6	0·781	0·797	0·814	0·831	0·847	0·881	0·914
+0·8	0·717	0·733	0·750	0·767	0·783	0·817	0·850
+1·0	0·657	0·674	0·690	0·707	0·724	0·757	0·790
+1·2	0·602	0·619	0·636	0·652	0·669	0·702	0·736
+1·4	0·552	0·569	0·585	0·602	0·619	0·652	0·685
+1·6	0·506	0·523	0·540	0·556	0·573	0·606	0·640
+1·8	0·465	0·482	0·498	0·515	0·532	0·565	0·598
+2·0	0·429	0·445	0·462	0·479	0·495	0·529	0·562

μ_{22}

x_3 \ $x_2{}^2$	0	0·25	0·50	0·75	1·00	1·50	2·00
−1·0	1·394	1·435	1·478	1·524	1·571	1·670	1·776
−0·8	1·295	1·329	1·366	1·404	1·445	1·532	1·627
−0·6	1·207	1·234	1·264	1·296	1·331	1·406	1·489
−0·4	1·129	1·150	1·173	1·199	1·227	1·290	1·363
−0·2	1·061	1·074	1·091	1·110	1·132	1·184	1·245
0	1·000	1·006	1·016	1·029	1·045	1·085	1·136
+0·2	0·946	0·946	0·949	0·955	0·965	0·994	1·035
+0·4	0·898	0·891	0·887	0·887	0·891	0·910	0·942
+0·6	0·855	0·841	0·831	0·825	0·823	0·832	0·855
+0·8	0·816	0·795	0·779	0·767	0·760	0·759	0·774
+1·0	0·781	0·753	0·731	0·714	0·702	0·692	0·700
+1·2	0·749	0·714	0·686	0·664	0·648	0·631	0·632
+1·4	0·718	0·678	0·645	0·619	0·598	0·575	0·571
+1·6	0·690	0·644	0·607	0·576	0·553	0·524	0·517
+1·8	0·662	0·612	0·571	0·538	0·512	0·480	0·471
+2·0	0·636	0·583	0·539	0·504	0·477	0·443	0·434

TABLE 4 (*Continued*)

μ_{33}

$\varkappa_2{}^2$ / $\varkappa_3$	0	0·25	0·50	0·75	1·00	1·50	2·00
—1·0	1·361	1·428	1·505	1·591	1·685	1·900	2·146
—0·8	1·265	1·313	1·370	1·436	1·510	1·684	1·891
—0·6	1·185	1·216	1·256	1·304	1·360	1·498	1·668
—0·4	1·116	1·135	1·160	1·192	1·233	1·337	1·473
—0·2	1·055	1·064	1·078	1·098	1·124	1·199	1·303
0	1·000	1·003	1·009	1·019	1·033	1·081	1·157
+0·2	0·948	0·949	0·949	0·951	0·957	0·981	1·031
+0·4	0·897	0·898	0·897	0·894	0·892	0·898	0·924
+0·6	0·847	0·851	0·850	0·845	0·838	0·828	0·833
+0·8	0·796	0·806	0·807	0·801	0·792	0·769	0·756
+1·0	0·745	0·761	0·766	0·762	0·752	0·720	0·691
+1·2	0·692	0·716	0·727	0·726	0·716	0·679	0·637
+1·4	0·639	0·671	0·688	0·692	0·684	0·643	0·592
+1·6	0·587	0·627	0·650	0·658	0·653	0·612	0·554
+1·8	0·537	0·583	0·612	0·625	0·623	0·584	0·524
+2·0	0·491	0·541	0·575	0·592	0·593	0·558	0·499

μ_{31}

$\varkappa_2{}^2$ / $\varkappa_3$	0	0·25	0·50	0·75	1·00	1·50	2·00
—1·0	—0·0638	—0·0710	—0·0781	—0·0850	—0·0918	—0·1052	—0·1181
—0·8	—0·0636	—0·0712	—0·0785	—0·0858	—0·0929	—0·1068	—0·1201
—0·6	—0·0638	—0·0717	—0·0794	—0·0870	—0·0944	—0·1088	—0·1227
—0·4	—0·0644	—0·0727	—0·0807	—0·0886	—0·0963	—0·1113	—0·1257
—0·2	—0·0653	—0·0740	—0·0824	—0·0907	—0·0987	—0·1143	—0·1292
0	—0·0667	—0·0757	—0·0846	—0·0932	—0·1016	—0·1177	—0·1331
+0·2	—0·0684	—0·0779	—0·0872	—0·0962	—0·1049	—0·1217	—0·1376
+0·4	—0·0705	—0·0805	—0·0902	—0·0996	—0·1087	—0·1262	—0·1426
+0·6	—0·0730	—0·0836	—0·0937	—0·1035	—0·1130	—0·1311	—0·1481
+0·8	—0·0760	—0·0870	—0·0976	—0·1079	—0·1177	—0·1364	—0·1539
+1·0	—0·0792	—0·0908	—0·1019	—0·1126	—0·1229	—0·1422	—0·1602
+1·2	—0·0829	—0·0950	—0·1066	—0·1177	—0·1283	—0·1483	—0·1667
+1·4	—0·0868	—0·0995	—0·1116	—0·1231	—0·1340	—0·1545	—0·1734
+1·6	—0·0908	—0·1041	—0·1167	—0·1286	—0·1399	—0·1609	—0·1800
+1·8	—0·0949	—0·1088	—0·1218	—0·1340	—0·1456	—0·1671	—0·1865
+2·0	—0·0989	—0·1132	—0·1267	—0·1393	—0·1511	—0·1729	—0·1925

P. G. GUEST

TABLE 4 (*Continued*)

μ_{21}

(Sign of μ_{21} opposite to that of x_2)

x_2^2 $\lvert x_2 \rvert$	0 0	0·25 0·500	0·50 0·707	0·75 0·866	1·00 1·000	1·50 1·225	2·00 1·414
x_3							
—1·0	0	0·0778	0·1092	0·1327	0·1521	0·1835	0·2089
—0·8	0	0·0814	0·1142	0·1387	0·1588	0·1914	0·2176
—0·6	0	0·0853	0·1195	0·1451	0·1660	0·1998	0·2267
—0·4	0	0·0895	0·1253	0·1519	0·1737	0·2086	0·2365
—0·2	0	0·0940	0·1314	0·1592	0·1818	0·2180	0·2467
0	0	0·0988	0·1379	0·1669	0·1905	0·2280	0·2575
+0·2	0	0·1039	0·1449	0·1752	0·1997	0·2385	0·2688
+0·4	0	0·1094	0·1524	0·1839	0·2094	0·2495	0·2806
+0·6	0	0·1152	0·1602	0·1931	0·2196	0·2610	0·2928
+0·8	0	0·1213	0·1685	0·2028	0·2302	0·2728	0·3053
+1·0	0	0·1278	0·1772	0·2129	0·2412	0·2850	0·3181
+1·2	0	0·1345	0·1861	0·2232	0·2525	0·2973	0·3308
+1·4	0	0·1414	0·1952	0·2336	0·2638	0·3095	0·3433
+1·6	0	0·1483	0·2043	0·2439	0·2748	0·3213	0·3552
+1·8	0	0·1550	0·2130	0·2538	0·2853	0·3323	0·3661
+2·0	0	0·1613	0·2211	0·2628	0·2949	0·3420	0·3755

μ_{32}

(Sign of μ_{32} opposite to that of x_2)

x_2^2 $\lvert x_2 \rvert$	0 0	0·25 0·500	0·50 0·707	0·75 0·866	1·00 1·000	1·50 1·225	2·00 1·414
x_3							
—1·0	0	0·1580	0·2202	0·2657	0·3024	0·3600	0·4044
—0·8	0	0·1648	0·2299	0·2776	0·3161	0·3764	0·4228
—0·6	0	0·1711	0·2389	0·2889	0·3292	0·3925	0·4410
—0·4	0	0·1766	0·2472	0·2994	0·3416	0·4081	0·4590
—0·2	0	0·1813	0·2544	0·3089	0·3531	0·4231	0·4767
0	0	0·1852	0·2606	0·3173	0·3636	0·4373	0·4940
+0·2	0	0·1880	0·2656	0·3245	0·3730	0·4507	0·5107
+0·4	0	0·1898	0·2694	0·3303	0·3811	0·4631	0·5270
+0·6	0	0·1905	0·2718	0·3348	0·3879	0·4746	0·5427
+0·8	0	0·1903	0·2729	0·3379	0·3933	0·4851	0·5578
+1·0	0	0·1891	0·2728	0·3397	0·3974	0·4944	0·5722
+1·2	0	0·1871	0·2716	0·3401	0·4002	0·5026	0·5857
+1·4	0	0·1845	0·2694	0·3395	0·4017	0·5096	0·5981
+1·6	0	0·1814	0·2666	0·3379	0·4022	0·5153	0·6089
+1·8	0	0·1783	0·2634	0·3358	0·4018	0·5196	0·6176
+2·0	0	0·1754	0·2604	0·3336	0·4010	0·5223	0·6233

It will be observed that the expressions given in Table 3 for $\Sigma\varkappa_{2k}{}^2 R_k$ and for $\Sigma\varkappa_{1k}\varkappa_{3k}R_k$ are the same.   This is a special case of the general equality

$$\sum_1^{3m}\varkappa_{jk}\varkappa_{mk}R_k= \sum_1^{3(m+q)} \varkappa_{j-q,k}\varkappa_{m+q,k}R_k, \quad \ldots\ldots\ldots\ldots (33)$$

which may be proved by induction.

TABLE 5

THE COEFFICIENTS $\alpha_{jk},\ \beta_{jk}$

$\alpha_{10}\quad 0$	$\beta_{10}=0$
$\alpha_{21}=n\mu_{21}$	$\beta_{21}=n\mu_{21}$
$\alpha_{20}\quad-n^2\mu_{11}/12$	$\beta_{20}=-n^2\mu_{11}/12$
$\alpha_{32}\quad-n\mu_{32}$	$\beta_{32}=n\mu_{32}$
$\alpha_{31}=n^2(\mu_{31}-\mu_{11}/12)$	$\beta_{31}=n^2(\mu_{31}+\mu_{32}\mu_{21}-\mu_{11}/12)$
$\alpha_{30}=n^3\mu_{21}\mu_{11}/12$	$\beta_{30}=n^3\mu_{11}(\mu_{21}-\mu_{32})/12$

The coefficients $\varkappa_{jk},\ \beta_{jk}$ may be determined by combining equations (20), (22), and (26).   The values so obtained are listed in Table 5 for values of j from 1 to 3.

V. The Standard Errors
(a) The Orthogonal Coefficients a_j

If the least squares polynomial is written in the form

$$u_p(\xi)=\sum_{j=0}^p a_j T_j(\xi), \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (34)$$

the standard error of the coefficient a_j is given by

$$\sigma^2(y)/\sigma^2(a_j)=\sum_\xi T_j{}^2(\xi),$$

where $\sigma(y)$ is the standard error of an observation.   Thus

$$\sigma^2(a_j)=\mu_{jj}{}^{-1}\{\sigma^2(y)/\Sigma T_j{}^2(\varepsilon)\}. \quad \ldots\ldots\ldots\ldots\ldots (35)$$

The values μ_{jj} are given in Table 4.   It will be seen that for all three values of j negative values of $\varkappa_3$ yield values for μ_{jj} greater than unity and positive values of $\varkappa_3$ yield values less than unity.   That is, the standard errors are reduced if the observations are crowded towards the extremes of the range and increased if they are crowded towards the centre of the range.

The effect of $\varkappa_2$ can be summarized as follows.   The standard error $\sigma(a_1)$ decreases only slowly as $\varkappa_2{}^2$ increases.   The standard error $\sigma(a_2)$ for negative values of $\varkappa_3$ decreases as $\varkappa_2{}^2$ increases, while for positive values of $\varkappa_3$ it increases as $\varkappa_2{}^2$ increases.   The standard error $\sigma(a_3)$ for negative values of $\varkappa_3$ decreases as $\varkappa_2{}^2$ increases, but varies only slowly for positive values of $\varkappa_3$.   Both $\sigma(a_2)$ and $\sigma(a_3)$ are almost independent of $\varkappa_2$ near $\varkappa_3=0$.

If the values $\varkappa_2,\ \varkappa_3$ likely to be encountered in practical examples are assumed to lie between -1 and $+1$, then the corresponding range of μ_{jj} is from $1\cdot685$ to $0\cdot657$.   Thus the range of variation of the standard errors, which are proportional to $\mu_{jj}{}^{-\frac12}$, is from $0\cdot77$ to $1\cdot23$, and the standard errors $\sigma(a_j)$ will lie within 25 per cent. of the values for the equally spaced case.

The coefficient a_j is identical with the coefficient b_{jj} in the power series expansion of the polynomial of degree j,

$$u_j(\xi) = \sum_{k=0}^{j} b_{jk}\xi^k. \quad \dots\dots\dots\dots\dots\dots (36)$$

The values of the coefficients a_j are independent of the choice of origin for the variable.

(b) The Fitted Values

The standard error of the fitted value $u_p(\xi)$ is given by the expression

$$\sigma^2[u_p(\xi)]/\sigma^2(y) = \sum_{j=0}^{p} \{T_j^2(\xi)/\sum_{\xi} T_j^2(\xi)\}$$

$$= n^{-1} \sum_{j=0}^{p} \{(\sum_{k=0}^{j} \beta_{jk}\xi^k)^2/\mu_{jj}R_j n^{2j}\}$$

$$= n^{-1} \sum_{j=0}^{p} \chi_{j0}, \text{ say.}$$

If the substitution $e = 2\xi/n$ is made, it is found that the quantities χ_{j0} are the following functions of e:

$$\chi_{00} = 1,$$

$$\chi_{10} = 3e^2/\mu_{11}, \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (37)$$

$$\chi_{20} = 1 \cdot 25(3e^2 + 6en^{-1}\beta_{21} + 12n^{-2}\beta_{20})^2/\mu_{22}, \quad \dots\dots\dots\dots (38)$$

$$\chi_{30} = 1 \cdot 75(5e^3 + 10e^2n^{-1}\beta_{32} + 20en^{-2}\beta_{31} + 40n^{-3}\beta_{30})^2/\mu_{33}. \dots (39)$$

The quantities $n^{-(j-k)}\beta_{jk}$ are functions of the coefficients μ, as given in Table 5. The standard error may be written in the form

$$\sigma[u_p(\xi)] = n^{-\frac{1}{2}}\rho_{p0}[e, \varkappa_2, \varkappa_3]\sigma(y). \quad \dots\dots\dots\dots (40)$$

For the case $p = 1$, the function ρ_{p0} is just $(1 + 3e^2/\mu_{11})^{\frac{1}{2}}$. Thus the standard error is a minimum at the value $\xi = 0$ (i.e. at $x = \Sigma x/n$), and increases symmetrically on each side of this value. If the variable $k = e\mu_{11}^{-\frac{1}{2}}$ is introduced, then $\rho_{10}^2 = 1 + 3k^2$ and the function $\rho_{10}(k)$ is identical with that tabulated in a previous paper (Guest 1950). When the value $\Sigma T_1^2(x)$ is known from the least squares calculations, k is given by

$$k = \left\{\frac{n}{3\Sigma T_1^2(x)}\right\}^{\frac{1}{2}}\left\{x - \frac{\Sigma x}{n}\right\},$$

and

$$\sigma[u_1(x)] = n^{-1}\rho_{10}(k)\sigma(y),$$

where $\rho_{10}(k)$ may be read off directly from the tables.

For the case $p = 2$, χ_{20} is roughly symmetrical about the value $e = -\mu_{21}$, and the major effect of the parameter $\varkappa_2$ is to make ρ_{20} roughly symmetrical about $e = \varkappa_2/5$. For this reason it is an advantage to introduce a new variable k, such that

$$k = e - \varkappa_2/5. \quad \dots\dots\dots\dots\dots\dots (41)$$

The function $\rho_{20}[k, \varkappa_2, \varkappa_3]$ is given in Table 6 for the range $k = -1 \cdot 4(0 \cdot 2) + 1 \cdot 4$, $\varkappa_2 = -1 \cdot 0(0 \cdot 5) + 1 \cdot 0$, $\varkappa_3 = -1 \cdot 0(0 \cdot 25) + 1 \cdot 0$.

Similarly, for the case $p=3$, it is an advantage to tabulate in terms of the variable k given by (41) rather than the variable e. The function $\rho_{30}[k, \varkappa_2, \varkappa_3]$ is given in Table 6 for the same range as the function ρ_{20}.

It is found that when $|\varkappa_2|$ is large the values of ρ_{20} and ρ_{30} near $|k|=0\cdot5$ are increased for points with $k\varkappa_2$ positive and decreased for points with $k\varkappa_2$ negative. The parameter $\varkappa_3$ has a much less marked effect. In general the values of ρ_{20} and ρ_{30} are increased when $\varkappa_3$ is positive and decreased when $\varkappa_3$ is negative.

The procedure for the rough estimation of the standard errors of the fitted values when the polynomial is of the second or third degree may be summarized as follows :

(i) Suppose the observations to be numbered, in order of x, by the values ε from $+\frac{1}{2}(n-1)$ to $-\frac{1}{2}(n-1)$. Write down the values for x_1, x_5 for $\varepsilon=\pm\frac{1}{2}(n-1)$; x_2, x_4 for $\varepsilon=\pm\frac{1}{4}(n-1)$; x_3 for $\varepsilon=0$, interpolating where necessary.

(ii) Calculate
$$\psi_2=2(x_1+x_5-2x_3),$$
$$\psi_3=2\cdot67\{(x_1-x_5)-2(x_2-x_4)\},$$
$$\psi_1=(x_1-x_5)+\psi_3/(n-1),$$
$$\varphi=(n-1)/\psi_1,$$
$$\varkappa_2=\{1+1/n\}\{\psi_2/\psi_1\},$$
$$\varkappa_3=\{1+2/n\}\{\psi_3/\psi_1\}.$$

(iii) Calculate
$$k=\frac{2\varphi}{n}x-\left(\frac{2\varphi}{n}\ \frac{\Sigma x}{n}+\frac{\varkappa_2}{5}\right),$$

for each value of x at which the standard error is required.

(iv) Find $\rho_{p0}(k, \varkappa'_2, \varkappa'_3)$ by interpolation in Table 6 for the values $\varkappa'_2, \varkappa'_3$ nearest to $\varkappa_2, \varkappa_3$.

(v) Then
$$\sigma[u_p(x)]\doteqdot n^{-\frac{1}{2}}\rho_{p0}(k, \varkappa'_2, \varkappa'_3)\sigma(y).$$

(c) The Polynomial Coefficients

Suppose that at the origin of the variable x the variable e has the value e_0. If the least squares polynomial is written as
$$u_p(x)=\sum_{j=0}^{p} c_{pj}x^j, \quad\dots\dots\dots\dots\dots (42)$$
then the standard error of the coefficient c_{pj} is given by an expression of the form
$$\sigma[c_{pj}(e_0,\varkappa_2,\varkappa_3)]=\varphi^j n^{-(j+\frac{1}{2})}\rho_{pj}(e_0,\varkappa_2,\varkappa_3)\sigma(y). \quad\dots (43)$$
The functions ρ_{pj} for $j=1$ are given by
$$\rho_{11}{}^2=12/\mu_{11}$$
$$\rho_{21}{}^2=\rho_{11}{}^2+180(e_0+n^{-1}\beta_{21})^2/\mu_{22},$$
$$\rho_{31}{}^2=\rho_{21}{}^2+175(3e_0{}^2+4e_0n^{-1}\beta_{32}+4n^{-2}\beta_{31})^2/\mu_{33}.$$
For $j=2$ the functions are given by
$$\rho_{22}{}^2=180/\mu_{22},$$
$$\rho_{32}{}^2=\rho_{22}{}^2+700(3e_0+2n^{-1}\beta_{32})^2/\mu_{33}.$$

B

TABLE 6

VALUES OF ρ_{20} AND ρ_{30}

x_3	$|x_2|$	$|k|$	kx_2 negative							ρ_{20}	kx_2 positive						
			1·4	1·2	1·0	0·8	0·6	0·4	0·2	0	0·2	0·4	0·6	0·8	1·0	1·2	1·4
$x_3=+1\cdot00$	$|x_2|=1\cdot0$	1·0	7·59	5·46	3·68	2·27	1·36	1·16	1·39	1·61	1·69	1·72	1·91	2·52	3·62	5·15	7·07
		0·5	7·54	5·51	3·81	2·48	1·57	1·20	1·26	1·38	1·44	1·51	1·81	2·55	3·71	5·26	7·16
		0	7·32	5·39	3·79	2·54	1·71	1·33	1·28	1·30	1·28	1·33	1·71	2·54	3·79	5·39	7·32
$x_3=+0\cdot75$	$|x_2|=1\cdot0$	1·0	7·02	5·03	3·36	2·07	1·30	1·19	1·42	1·62	1·68	1·68	1·85	2·43	3·48	4·96	6·79
		0·5	7·11	5·17	3·56	2·30	1·48	1·21	1·30	1·42	1·45	1·48	1·74	2·43	3·55	5·06	6·90
		0	7·01	5·14	3·59	2·40	1·63	1·32	1·32	1·35	1·32	1·32	1·63	2·40	3·59	5·14	7·01
$x_3=+0\cdot50$	$|x_2|=1\cdot0$	1·0	6·49	4·62	3·07	1·90	1·25	1·23	1·46	1·63	1·66	1·64	1·78	2·33	3·34	4·75	6·52
		0·5	6·70	4·85	3·31	2·13	1·41	1·23	1·35	1·46	1·47	1·47	1·68	2·31	3·39	4·84	6·63
		0	6·70	4·88	3·39	2·26	1·56	1·32	1·35	1·39	1·35	1·32	1·56	2·26	3·39	4·88	6·70
$x_3=+0\cdot25$	$|x_2|=1\cdot0$	1·0	6·00	4·24	2·80	1·74	1·22	1·27	1·50	1·64	1·66	1·62	1·73	2·23	3·19	4·55	6·24
		0·5	6·30	4·53	3·07	1·98	1·36	1·26	1·40	1·50	1·50	1·46	1·62	2·20	3·22	4·62	6·35
		0	6·38	4·63	3·19	2·12	1·50	1·33	1·40	1·45	1·40	1·33	1·50	2·12	3·19	4·63	6·38
$x_3=0$	$|x_2|=1\cdot0$	1·0	5·54	3·89	2·56	1·61	1·21	1·32	1·54	1·66	1·66	1·60	1·67	2·14	3·05	4·34	5·97
		0·5	5·91	4·22	2·85	1·84	1·32	1·29	1·45	1·55	1·53	1·46	1·56	2·09	3·06	4·40	6·07
		0	6·05	4·37	3·00	1·99	1·44	1·35	1·44	1·50	1·44	1·35	1·44	1·99	3·00	4·37	6·05
$x_3=-0\cdot25$	$|x_2|=1\cdot0$	1·0	5·11	3·57	2·34	1·50	1·22	1·36	1·58	1·68	1·66	1·58	1·63	2·04	2·90	4·14	5·69
		0·5	5·52	3·92	2·63	1·71	1·30	1·33	1·50	1·59	1·56	1·47	1·52	1·98	2·89	4·18	5·77
		0	5·73	4·11	2·81	1·87	1·41	1·37	1·49	1·55	1·49	1·37	1·41	1·87	2·81	4·11	5·73
$x_3=-0\cdot50$	$|x_2|=1\cdot0$	1·0	4·71	3·27	2·14	1·42	1·24	1·41	1·61	1·71	1·67	1·58	1·59	1·95	2·76	3·93	5·41
		0·5	5·15	3·64	2·43	1·61	1·29	1·38	1·55	1·63	1·59	1·48	1·49	1·88	2·73	3·95	5·48
		0	5·40	3·86	2·62	1·76	1·38	1·40	1·54	1·61	1·54	1·40	1·38	1·76	2·62	3·86	5·40
$x_3=-0\cdot75$	$|x_2|=1\cdot0$	1·0	4·33	2·99	1·96	1·35	1·26	1·46	1·65	1·73	1·69	1·58	1·56	1·87	2·62	3·72	5·13
		0·5	4·80	3·37	2·24	1·52	1·29	1·42	1·60	1·68	1·63	1·50	1·46	1·79	2·57	3·72	5·18
		0	5·07	3·61	2·45	1·67	1·37	1·44	1·59	1·66	1·59	1·44	1·37	1·67	2·45	3·61	5·07
$x_3=-1\cdot00$	$|x_2|=1\cdot0$	1·0	3·99	2·74	1·81	1·31	1·30	1·51	1·69	1·76	1·70	1·58	1·53	1·79	2·48	3·52	4·86
		0·5	4·45	3·11	2·07	1·45	1·31	1·47	1·65	1·72	1·66	1·52	1·45	1·70	2·41	3·50	4·88
		0	4·75	3·36	2·28	1·58	1·37	1·48	1·64	1·70	1·64	1·48	1·37	1·58			

x_3	$\lvert x_2\rvert$	$\lvert k\rvert$	1·4	1·2	1·0	0·8	0·6	0·4	0·2	0	0·2	0·4	0·6	0·8	1·0	1·2	1·4
$x_3=+1\cdot00$	1·0	1·0	14·0	7·83	4·06	2·58	2·45	2·32	1·95	1·61	1·50	1·41	1·38	2·81	6·27	11·8	19·7
	0·5	0·5	15·5	8·90	4·61	2·55	2·18	2·07	1·73	1·39	1·37	1·48	1·62	2·75	5·74	10·8	18·1
	0	0	16·9	9·91	5·20	2·62	1·89	1·76	1·49	1·30	1·49	1·76	1·89	2·62	5·20	9·91	16·9
$x_3=+0\cdot75$	1·0	1·0	13·5	7·54	3·89	2·48	2·38	2·26	1·91	1·62	1·56	1·50	1·35	2·49	5·71	11·0	18·5
	0·5	0·5	14·9	8·51	4·39	2·43	2·10	2·01	1·71	1·42	1·43	1·52	1·56	2·52	5·37	10·2	17·3
	0	0	16·2	9·42	4·91	2·46	1·82	1·74	1·51	1·35	1·51	1·74	1·82	2·46	4·91	9·42	16·2
$x_3=+0\cdot50$	1·0	1·0	13·0	7·23	3·72	2·39	2·31	2·20	1·88	1·63	1·62	1·59	1·36	2·17	5·14	10·1	17·3
	0·5	0·5	14·2	8·12	4·17	2·31	2·04	1·97	1·70	1·46	1·49	1·57	1·53	2·30	4·98	9·63	16·4
	0	0	15·4	8·94	4·61	2·31	1·77	1·74	1·54	1·39	1·54	1·74	1·77	2·31	4·61	8·94	15·4
$x_3=+0\cdot25$	1·0	1·0	12·4	6·92	3·55	2·30	2·25	2·15	1·85	1·64	1·68	1·69	1·42	1·89	4·55	9·19	15·9
	0·5	0·5	13·6	7·74	3·94	2·20	1·99	1·95	1·70	1·50	1·56	1·63	1·53	2·09	4·58	9·02	15·5
	0	0	14·7	8·46	4·31	2·15	1·74	1·75	1·58	1·45	1·58	1·75	1·74	2·15	4·31	8·46	14·7
$x_3=0$	1·0	1·0	11·9	6·60	3·37	2·21	2·20	2·12	1·84	1·66	1·73	1·77	1·50	1·66	3·97	8·26	14·6
	0·5	0·5	13·0	7·34	3·70	2·09	1·96	1·94	1·71	1·55	1·62	1·71	1·56	1·89	4·15	8·37	14·6
	0	0	14·0	7·97	4·00	2·01	1·73	1·78	1·62	1·50	1·62	1·78	1·73	2·01	4·00	7·97	14·0
$x_3=-0\cdot25$	1·0	1·0	11·3	6·26	3·19	2·13	2·15	2·09	1·84	1·68	1·78	1·85	1·60	1·51	3·41	7·32	13·2
	0·5	0·5	12·4	6·94	3·46	1·99	1·94	1·94	1·74	1·59	1·69	1·78	1·61	1·73	3·73	7·70	13·6
	0	0	13·2	7·47	3·68	1·88	1·74	1·82	1·67	1·55	1·67	1·82	1·74	1·88	3·68	7·47	13·2
$x_3=-0\cdot50$	1·0	1·0	10·7	5·90	3·01	2·06	2·12	2·08	1·85	1·71	1·81	1·92	1·71	1·43	2·89	6·41	11·8
	0·5	0·5	11·8	6·52	3·22	1·91	1·94	1·96	1·76	1·64	1·74	1·86	1·69	1·61	3·30	7·00	12·6
	0	0	12·5	6·94	3·36	1·77	1·78	1·87	1·72	1·61	1·72	1·87	1·78	1·77	3·36	6·94	12·5
$x_3=-0\cdot75$	1·0	1·0	10·1	5·54	2·82	1·99	2·09	2·07	1·86	1·73	1·84	1·97	1·80	1·43	2·44	5·55	10·4
	0·5	0·5	11·1	6·09	2·97	1·84	1·95	1·98	1·80	1·68	1·80	1·93	1·77	1·54	2·89	6·29	11·5
	0	0	11·6	6·39	3·03	1·69	1·82	1·93	1·77	1·66	1·77	1·93	1·82	1·69	3·03	6·39	11·6
$x_3=-1\cdot00$	1·0	1·0	9·46	5·17	2·64	1·94	2·08	2·07	1·88	1·76	1·86	2·00	1·88	1·48	2·06	4·75	9·15
	0·5	0·5	10·4	5·63	2·72	1·79	1·97	2·01	1·83	1·72	1·84	1·99	1·86	1·53	2·50	5·57	10·5
	0	0	10·8	5·83	2·71	1·65	1·88	1·98	1·82	1·70	1·82	1·98	1·88	1·65	2·71	5·83	10·8

VI. Illustrative Examples
(a) Exact Calculations

(i) Example 1

This example was used in a previous paper (Guest 1952). There are 16 observations, at the following values of x:

$$+7 \cdot 9 \quad +7 \cdot 2 \quad +6 \cdot 25 \quad +6 \cdot 1 \quad +4 \cdot 8 \quad +3 \cdot 5 \quad +2 \cdot 8 \quad +0 \cdot 8$$
$$-1 \cdot 25 \quad -2 \cdot 2 \quad -3 \cdot 65 \quad -4 \cdot 4 \quad -5 \cdot 5 \quad -6 \cdot 4 \quad -7 \cdot 3 \quad -9 \cdot 0.$$

The parameters are

$$\varphi = 0 \cdot 95576, \; \varkappa_2 = -0 \cdot 1227, \; \varkappa_3 = -0 \cdot 6415.$$

For the coefficients μ_{jj} interpolation in Table 4 gives

$$\mu_{11} = 1 \cdot 282 \; (1 \cdot 287), \; \mu_{22} = 1 \cdot 227 \; (1 \cdot 229), \; \mu_{33} = 1 \cdot 204 \; (1 \cdot 585).$$

The true values, obtained from the calculation of $\varphi^{2j} \Sigma T_j^2(x) / \Sigma T_j^2(\varepsilon)$, are shown in brackets.

Interpolation in Table 4 gives for the coefficients μ_{jk}

$$\mu_{21} = +0 \cdot 0207, \; \mu_{32} = +0 \cdot 0417, \; \mu_{33} = -0 \cdot 0643.$$

Application of formulae (37) to (40), with $e = 0 \cdot 11947x + 0 \cdot 0026$, leads to the functions ρ_{20}, ρ_{30} tabulated below, the true values being shown in brackets.

x	ρ_{20}	ρ_{30}
-10	$3 \cdot 52 \; (3 \cdot 55)$	$5 \cdot 8 \;\; (5 \cdot 2)$
$- 7 \cdot 5$	$1 \cdot 96 \; (1 \cdot 99)$	$2 \cdot 0 \;\; (2 \cdot 0)$
$- 5$	$1 \cdot 38 \; (1 \cdot 38)$	$1 \cdot 87 \; (1 \cdot 73)$
$- 2 \cdot 5$	$1 \cdot 53 \; (1 \cdot 54)$	$1 \cdot 83 \; (1 \cdot 69)$
0	$1 \cdot 64 \; (1 \cdot 65)$	$1 \cdot 64 \; (1 \cdot 68)$
$+ 2 \cdot 5$	$1 \cdot 48 \; (1 \cdot 48)$	$1 \cdot 86 \; (1 \cdot 89)$
$+ 5$	$1 \cdot 36 \; (1 \cdot 36)$	$1 \cdot 74 \; (1 \cdot 68)$
$+ 7 \cdot 5$	$2 \cdot 12 \; (2 \cdot 11)$	$2 \cdot 3 \;\; (2 \cdot 4)$
$+10$	$3 \cdot 84 \; (3 \cdot 84)$	$7 \cdot 2 \;\; (7 \cdot 1)$

(ii) Example 2

The 67 observations are those of Jaeger and von Steinwehr (1921), obtained in their determination of the mechanical equivalent of heat. Proceeding as in Example 1, the following values are obtained.

x : $+29 \cdot 60 \quad +28 \cdot 36 \quad +26 \cdot 96 \quad +25 \cdot 79 \quad +25 \cdot 56 \quad +24 \cdot 34 \quad +23 \cdot 09 \quad +19 \cdot 41$
$+17 \cdot 19 \quad +16 \cdot 64 \quad +15 \cdot 79 \quad +15 \cdot 75 \quad +14 \cdot 39 \quad +14 \cdot 32 \quad +13 \cdot 98 \quad +12 \cdot 54$
$+11 \cdot 49 \quad +10 \cdot 60 \quad + 9 \cdot 19 \quad + 9 \cdot 15 \quad + 7 \cdot 76 \quad + 6 \cdot 33 \quad + 6 \cdot 24 \quad + 5 \cdot 55$
$+ 5 \cdot 36 \quad + 4 \cdot 82 \quad + 4 \cdot 11 \quad + 3 \cdot 96 \quad + 3 \cdot 24 \quad + 2 \cdot 53 \quad + 1 \cdot 80 \quad + 1 \cdot 41$
$+ 1 \cdot 15 \quad + 1 \cdot 13 \quad + 0 \cdot 01 \quad - 0 \cdot 05 \quad - 0 \cdot 25 \quad - 1 \cdot 17 \quad - 1 \cdot 18 \quad - 1 \cdot 45$
$- 1 \cdot 58 \quad - 1 \cdot 58 \quad - 2 \cdot 59 \quad - 2 \cdot 67 \quad - 3 \cdot 61 \quad - 3 \cdot 69 \quad - 5 \cdot 01 \quad - 5 \cdot 12$
$- 6 \cdot 00 \quad - 6 \cdot 03 \quad - 6 \cdot 97 \quad - 7 \cdot 13 \quad - 7 \cdot 40 \quad - 7 \cdot 41 \quad - 7 \cdot 94 \quad - 8 \cdot 39$
$- 8 \cdot 55 \quad - 8 \cdot 67 \quad - 9 \cdot 98 \quad -11 \cdot 05 \quad -11 \cdot 19 \quad -11 \cdot 49 \quad -12 \cdot 59 \quad -13 \cdot 62$
$-13 \cdot 85 \quad -14 \cdot 35 \quad -15 \cdot 25.$

$$\varphi = 1 \cdot 43378, \quad \varkappa_2 = 0 \cdot 5871, \quad \varkappa_3 = 0 \cdot 7403.$$

$$\mu_{11} = 0 \cdot 759 \, (0 \cdot 760), \quad \mu_{22} = 0 \cdot 803 \, (0 \cdot 810), \quad \mu_{33} = 0 \cdot 820 \, (0 \cdot 770).$$

$$\mu_{21} = -0 \cdot 1391, \quad \mu_{32} = -0 \cdot 2250, \quad \mu_{31} = -0 \cdot 0899.$$

$$e = 0 \cdot 042799x - 0 \cdot 129.$$

x	ρ_{20}	ρ_{30}
-20	$4 \cdot 27 \, (4 \cdot 31)$	$7 \cdot 6 \quad (7 \cdot 8)$
-15	$2 \cdot 77 \, (2 \cdot 77)$	$3 \cdot 55 \, (3 \cdot 61)$
-10	$1 \cdot 70 \, (1 \cdot 69)$	$1 \cdot 70 \, (1 \cdot 69)$
$-\ 5$	$1 \cdot 22 \, (1 \cdot 22)$	$1 \cdot 49 \, (1 \cdot 51)$
0	$1 \cdot 28 \, (1 \cdot 27)$	$1 \cdot 46 \, (1 \cdot 47)$
$+\ 5$	$1 \cdot 43 \, (1 \cdot 43)$	$1 \cdot 43 \, (1 \cdot 43)$
$+10$	$1 \cdot 49 \, (1 \cdot 49)$	$1 \cdot 71 \, (1 \cdot 73)$
$+15$	$1 \cdot 51 \, (1 \cdot 51)$	$2 \cdot 05 \, (2 \cdot 10)$
$+20$	$1 \cdot 79 \, (1 \cdot 78)$	$2 \cdot 15 \, (2 \cdot 18)$
$+25$	$2 \cdot 53 \, (2 \cdot 52)$	$2 \cdot 53 \, (2 \cdot 53)$
$+30$	$3 \cdot 76 \, (3 \cdot 77)$	$4 \cdot 8 \quad (4 \cdot 9)$

(iii) *Example 3*

This is an example used by Kendall (1948).

$$x: \ +101 \quad +87 \quad +76 \quad +62 \quad +53 \quad +41 \quad +29 \quad +13$$
$$+\ 3 \quad -\ 6 \quad -18 \quad -29 \quad -35 \quad -40 \quad -45 \quad -50.$$

$$\varphi = 0 \cdot 103080, \quad \varkappa_2 = +0 \cdot 4706, \quad \varkappa_3 = -0 \cdot 3920.$$

$$\mu_{11} = 1 \cdot 181 \, (1 \cdot 181), \quad \mu_{22} = 1 \cdot 145 \, (1 \cdot 155), \quad \mu_{33} = 1 \cdot 130 \, (1 \cdot 232).$$

$$\mu_{21} = -0 \cdot 0844, \quad \mu_{32} = -0 \cdot 1664, \quad \mu_{31} = -0 \cdot 0718.$$

$$e = 0 \cdot 012885x - 0 \cdot 195.$$

x	ρ_{20}	ρ_{30}
$-\ 60$	$2 \cdot 89 \, (2 \cdot 89)$	$4 \cdot 51 \, (4 \cdot 52)$
$-\ 40$	$1 \cdot 67 \, (1 \cdot 68)$	$1 \cdot 68 \, (1 \cdot 68)$
$-\ 20$	$1 \cdot 28 \, (1 \cdot 28)$	$1 \cdot 73 \, (1 \cdot 73)$
0	$1 \cdot 45 \, (1 \cdot 46)$	$1 \cdot 79 \, (1 \cdot 80)$
$+\ 20$	$1 \cdot 61 \, (1 \cdot 61)$	$1 \cdot 62 \, (1 \cdot 62)$
$+\ 40$	$1 \cdot 56 \, (1 \cdot 56)$	$1 \cdot 78 \, (1 \cdot 74)$
$+\ 60$	$1 \cdot 45 \, (1 \cdot 44)$	$1 \cdot 97 \, (1 \cdot 93)$
$+\ 80$	$1 \cdot 75 \, (1 \cdot 75)$	$1 \cdot 87 \, (1 \cdot 87)$
$+100$	$2 \cdot 77 \, (2 \cdot 81)$	$3 \cdot 28 \, (3 \cdot 30)$
$+120$	$4 \cdot 45 \, (4 \cdot 49)$	$8 \cdot 0 \quad (7 \cdot 9)$

(b) *Approximate Calculations*

In practical examples it would require too much time to undertake a least squares determination of the parameters φ, $\varkappa_2$, $\varkappa_3$. The approximate scheme described in Section V (*b*) based on formulae (**11**), (**12**), and (**13**), will now be applied to the three examples treated in Section VI (*a*).

(i) *Example 1*

$$x_1 = +7 \cdot 9, \quad x_3 = -0 \cdot 225, \quad x_2 = +5 \cdot 1,$$
$$x_5 = -9 \cdot 0, \qquad\qquad\qquad x_4 = -4 \cdot 7.$$

$$\psi_2 = 2\{x_1 + x_5 - 2x_3\} = -1 \cdot 30,$$
$$\psi_3 = \frac{8}{3}\{(x_1 - x_5) - 2(x_2 - x_4)\} = -7 \cdot 2,$$

$$\psi_1 = (x_1 - x_5) + \psi_3/(n-1) = 16 \cdot 42.$$
$$\varphi = (n-1)/\psi_1 = 0 \cdot 914 \ (0 \cdot 956),$$
$$\varkappa_2 = \{1 + 1/n\}\{\psi_2/\psi_1\} = -0 \cdot 084 \ (-0 \cdot 123),$$
$$\varkappa_3 = \{1 + 2/n\}\{\psi_3/\psi_1\} = -0 \cdot 499 \ (-0 \cdot 641).$$

Interpolation in Table 4 gives

$$\mu_{11} = 1 \cdot 22, \quad \mu_{22} = 1 \cdot 17, \quad \mu_{33} = 1 \cdot 15,$$

and hence the values $\Sigma T_j{}^2(x) = \varphi^{-2j}\mu_{jj}\Sigma T_j{}^2(\varepsilon)$ are $\Sigma T_1{}^2(x) = 497 \ (479)$, $\Sigma T_2{}^2(x) = 958 \times 10 \ (841)$, $\Sigma T_3{}^2(x) = 179 \times 10^3 \ (189)$.

On interpolating in Table 6, with $\varkappa_2 = 0$ and $\varkappa_3 = -0 \cdot 5$, and using (**41**), the following values are obtained.

$$k = 0 \cdot 1142x + 0 \cdot 019.$$

x	k	ρ_{20}	ρ_{30}
-10	$-1 \cdot 12$	$3 \cdot 4 \quad (3 \cdot 6)$	$5 \cdot 5 \quad (5 \cdot 2)$
$-7 \cdot 5$	$-0 \cdot 84$	$1 \cdot 93 \ (1 \cdot 99)$	$2 \cdot 1 \quad (2 \cdot 0)$
-5	$-0 \cdot 55$	$1 \cdot 38 \ (1 \cdot 38)$	$1 \cdot 80 \ (1 \cdot 73)$
$-2 \cdot 5$	$-0 \cdot 27$	$1 \cdot 49 \ (1 \cdot 54)$	$1 \cdot 77 \ (1 \cdot 69)$
0	$+0 \cdot 02$	$1 \cdot 60 \ (1 \cdot 65)$	$1 \cdot 62 \ (1 \cdot 68)$
$+2 \cdot 5$	$+0 \cdot 30$	$1 \cdot 47 \ (1 \cdot 48)$	$1 \cdot 80 \ (1 \cdot 89)$
$+5$	$+0 \cdot 59$	$1 \cdot 38 \ (1 \cdot 36)$	$1 \cdot 78 \ (1 \cdot 68)$
$+7 \cdot 5$	$+0 \cdot 88$	$2 \cdot 10 \ (2 \cdot 11)$	$2 \cdot 4 \quad (2 \cdot 4)$
$+10$	$+1 \cdot 16$	$3 \cdot 6 \quad (3 \cdot 8)$	$6 \cdot 2 \quad (7 \cdot 1)$

(ii) *Example 2*

$$x_1 = +29 \cdot 60, \ x_5 = -15 \cdot 25, \ x_3 = +1 \cdot 13, \ x_2 = +11 \cdot 05, \ x_4 = -6 \cdot 50.$$
$$\psi_2 = +24 \cdot 18, \ \psi_3 = +26 \cdot 00, \ \psi_1 = +45 \cdot 24.$$
$$\varphi = 1 \cdot 459 \ (1 \cdot 434), \ \varkappa_2 = +0 \cdot 543 \ (0 \cdot 587), \ \varkappa_3 = +0 \cdot 592 \ (0 \cdot 740).$$
$$\mu_{11} = 0 \cdot 80, \ \mu_{22} = 0 \cdot 84, \ \mu_{33} = 0 \cdot 85.$$
$$\Sigma T_1{}^2(x) = 94 \times 10^2 \ (93), \ \Sigma T_2{}^2(x) = 139 \times 10^4 \ (144), \ \Sigma T_3{}^2(x) = 191 \times 10^6 \ (191).$$
$$k = 0 \cdot 0436x - 0 \cdot 240.$$

x	k	ρ_{20}	ρ_{30}
-20	$-1 \cdot 11$	$4 \cdot 2 \quad (4 \cdot 3)$	$7 \cdot 5 \quad (7 \cdot 8)$
-15	$-0 \cdot 89$	$2 \cdot 66 \ (2 \cdot 77)$	$3 \cdot 5 \quad (3 \cdot 6)$
-10	$-0 \cdot 68$	$1 \cdot 70 \ (1 \cdot 69)$	$1 \cdot 84 \ (1 \cdot 69)$
-5	$-0 \cdot 46$	$1 \cdot 28 \ (1 \cdot 22)$	$1 \cdot 56 \ (1 \cdot 51)$
0	$-0 \cdot 24$	$1 \cdot 32 \ (1 \cdot 27)$	$1 \cdot 51 \ (1 \cdot 47)$
$+5$	$-0 \cdot 02$	$1 \cdot 45 \ (1 \cdot 43)$	$1 \cdot 46 \ (1 \cdot 43)$
$+10$	$+0 \cdot 20$	$1 \cdot 47 \ (1 \cdot 49)$	$1 \cdot 70 \ (1 \cdot 73)$
$+15$	$+0 \cdot 41$	$1 \cdot 48 \ (1 \cdot 51)$	$1 \cdot 97 \ (2 \cdot 10)$
$+20$	$+0 \cdot 63$	$1 \cdot 77 \ (1 \cdot 78)$	$2 \cdot 08 \ (2 \cdot 18)$
$+25$	$+0 \cdot 85$	$2 \cdot 58 \ (2 \cdot 52)$	$2 \cdot 8 \quad (2 \cdot 5)$
$+30$	$+1 \cdot 07$	$3 \cdot 9 \quad (3 \cdot 8)$	$5 \cdot 6 \quad (4 \cdot 9)$

(iii) *Example 3*

$$x_1 = +101, \quad x_5 = -50, \quad x_3 = +8, \quad x_2 = +55, \quad x_4 = -30 \cdot 5.$$
$$\psi_2 = +70, \quad \psi_3 = -53 \cdot 3, \quad \psi_1 = 147 \cdot 45.$$
$$\varphi = 0 \cdot 1017 \; (0 \cdot 1031), \quad \varkappa_2 = +0 \cdot 506 \; (0 \cdot 471), \quad \varkappa_3 = -0 \cdot 411 \; (0 \cdot 392).$$
$$\mu_{11} = 1 \cdot 19, \quad \mu_{22} = 1 \cdot 15, \quad \mu_{33} = 1 \cdot 14.$$
$$\Sigma T_1{}^2(x) = 391 \times 10^2 \; (378), \quad \Sigma T_2{}^2(x) = 614 \times 10^5 \; (585), \quad \Sigma T_3{}^2(x) = 93 \times 10^9 \; (93).$$
$$k = 0 \cdot 01271x - 0 \cdot 293.$$

x	k	ρ_{20}	ρ_{30}
$- 60$	$-1 \cdot 06$	$2 \cdot 8 \;\; (2 \cdot 9)$	$4 \cdot 4 \;\; (4 \cdot 5)$
$- 40$	$-0 \cdot 80$	$1 \cdot 61 \; (1 \cdot 68)$	$1 \cdot 61 \; (1 \cdot 68)$
$- 20$	$-0 \cdot 55$	$1 \cdot 31 \; (1 \cdot 28)$	$1 \cdot 73 \; (1 \cdot 73)$
0	$-0 \cdot 29$	$1 \cdot 47 \; (1 \cdot 46)$	$1 \cdot 79 \; (1 \cdot 80)$
$+ 20$	$-0 \cdot 04$	$1 \cdot 61 \; (1 \cdot 61)$	$1 \cdot 66 \; (1 \cdot 62)$
$+ 40$	$+0 \cdot 22$	$1 \cdot 58 \; (1 \cdot 56)$	$1 \cdot 78 \; (1 \cdot 74)$
$+ 60$	$+0 \cdot 47$	$1 \cdot 48 \; (1 \cdot 44)$	$1 \cdot 95 \; (1 \cdot 93)$
$+ 80$	$+0 \cdot 72$	$1 \cdot 72 \; (1 \cdot 75)$	$1 \cdot 92 \; (1 \cdot 87)$
$+100$	$+0 \cdot 98$	$2 \cdot 64 \; (2 \cdot 81)$	$3 \cdot 1 \;\; (3 \cdot 3)$
$+120$	$+1 \cdot 23$	$4 \cdot 2 \;\; (4 \cdot 5)$	$7 \cdot 3 \;\; (7 \cdot 9)$

VII. Deviations of the Errors from the Exact Least Squares Values

It is necessary to examine whether the smoothed set of points $X(\varepsilon)$ will be an adequate substitute for the actual set $x(\varepsilon)$ in the calculation of the standard errors. The departures $x(\varepsilon) - X(\varepsilon)$ may be represented by additional terms $k_4 T_4(\varepsilon) + k_5 T_5(\varepsilon)$. Thus to obtain a closer approximation to the observed points the expression $\xi(\varepsilon)$ can be written in the form

$$\xi(\varepsilon) = \xi_3(\varepsilon) + 4\varkappa_4 \tau_4(\varepsilon) + 8\varkappa_5 \tau_5(\varepsilon), \quad \dots\dots\dots (44)$$

where $\xi_3(\varepsilon)$ is the expression given in equation (4).

If the function $\xi(\varepsilon)$ given by (44) is to represent the points of observation, it must satisfy the condition $\Delta \xi(\varepsilon) \geqslant 0$. Thus it is required to find, for a given $\xi_3(\varepsilon)$, the ranges of the parameters $\varkappa_4$, $\varkappa_5$ for which this condition is satisfied. The ranges of these parameters and the effect they have on μ_{jj} will depend to some extent on the value of n. However, since no great accuracy is required, it is simplest to take some definite value of n, say 50, and make use of the tabulated values of the orthogonal polynomials for this number of observations. A table of finite difference values for the polynomials can then be drawn up and used to estimate the allowable ranges of $\varkappa_4$, $\varkappa_5$.

Table 7 gives the finite difference values for $n = 50$. As an illustration of the method of procedure, the calculation of the maximum positive value of $\varkappa_5$ when $\varkappa_2 = \varkappa_3 = 0$ will be outlined. The finite difference is

$$\Delta \xi(\varepsilon) = 1 + \varkappa_4 \{4 \Delta \tau_4(\varepsilon)\} + \varkappa_5 \{8 \Delta \tau_5(\varepsilon)\} \geqslant 0,$$

and so

$$\varkappa_5 \{8 \Delta \tau_5(\varepsilon)\} \geqslant -1 - \varkappa_4 \{4 \Delta \tau_4(\varepsilon)\},$$

and, since $\Delta \tau_5(\varepsilon)$ is an even function of ε, $\Delta \tau_4(\varepsilon)$ is an odd function of ε,

$$\varkappa_5 \{8 \Delta \tau_5(\varepsilon)\} \geqslant -1 + | \varkappa_4 \{4 \Delta \tau_4(\varepsilon)\} |.$$

Now, from Table 7, the minimum value of $8\Delta\tau_5(\varepsilon)$ s $-0\cdot156$. Hence the maximum positive value of $\varkappa_5$ occurs when $\varkappa_4=0$, and has the value $1/0\cdot156=6\cdot4$.

Examination of the table of finite differences leads in a similar manner to the following conclusions :

(a) When $\xi_3(\varepsilon)=\varepsilon$ (i.e. when $\varkappa_2=\varkappa_3=0$), values of $\varkappa_5$ from $-1\cdot4$ to $+6\cdot4$ can occur. The maximum value of $|\varkappa_4|$ depends on the value of $\varkappa_5$, being zero at the extreme values of $\varkappa_5$ and reaching a maximum of 3 when $\varkappa_5=+3$.

(b) The only major effect of the parameter $\varkappa_3$ is to permit larger negative values of $\varkappa_5$ (maximum negative $\varkappa_5=-3\cdot0$ when $\varkappa_3=+1\cdot3$).

TABLE 7

FINITE DIFFERENCES AND VALUES OF μ'_{jj} FOR $n=50$

		Finite Differences		
ε	$\Delta\tau_2$	$2\Delta\tau_3-\dfrac{1}{5}\Delta\tau_1$	$4\Delta\tau_4$	$8\Delta\tau_5$
$23\frac{1}{2}$	$0\cdot96$	$0\cdot883$	$0\cdot948$	$0\cdot711$
$22\frac{1}{2}$	$0\cdot92$	$0\cdot770$	$0\cdot771$	$0\cdot504$
$21\frac{1}{2}$	$0\cdot88$	$0\cdot662$	$0\cdot611$	$0\cdot332$
$20\frac{1}{2}$	$0\cdot84$	$0\cdot559$	$0\cdot467$	$0\cdot192$
$19\frac{1}{2}$	$0\cdot80$	$0\cdot461$	$0\cdot340$	$0\cdot079$
$18\frac{1}{2}$	$0\cdot76$	$0\cdot367$	$0\cdot228$	$-0\cdot006$
$17\frac{1}{2}$	$0\cdot72$	$0\cdot278$	$0\cdot131$	$-0\cdot071$
$16\frac{1}{2}$	$0\cdot68$	$0\cdot194$	$0\cdot048$	$-0\cdot115$
$15\frac{1}{2}$	$0\cdot64$	$0\cdot115$	$-0\cdot023$	$-0\cdot142$
$14\frac{1}{2}$	$0\cdot60$	$0\cdot040$	$-0\cdot081$	$-0\cdot156$
$13\frac{1}{2}$	$0\cdot56$	$-0\cdot029$	$-0\cdot127$	$-0\cdot156$
$12\frac{1}{2}$	$0\cdot52$	$-0\cdot094$	$-0\cdot163$	$-0\cdot148$
$11\frac{1}{2}$	$0\cdot48$	$-0\cdot154$	$-0\cdot190$	$-0\cdot132$
$10\frac{1}{2}$	$0\cdot44$	$-0\cdot209$	$-0\cdot205$	$-0\cdot110$
$9\frac{1}{2}$	$0\cdot40$	$-0\cdot260$	$-0\cdot214$	$-0\cdot083$
$8\frac{1}{2}$	$0\cdot36$	$-0\cdot305$	$-0\cdot215$	$-0\cdot055$
$7\frac{1}{2}$	$0\cdot32$	$-0\cdot346$	$-0\cdot208$	$-0\cdot026$
$6\frac{1}{2}$	$0\cdot28$	$-0\cdot382$	$-0\cdot195$	$0\cdot004$
$5\frac{1}{2}$	$0\cdot24$	$-0\cdot413$	$-0\cdot178$	$0\cdot030$
$4\frac{1}{2}$	$0\cdot20$	$-0\cdot439$	$-0\cdot155$	$0\cdot056$
$3\frac{1}{2}$	$0\cdot16$	$-0\cdot461$	$-0\cdot128$	$0\cdot077$
$2\frac{1}{2}$	$0\cdot12$	$-0\cdot478$	$-0\cdot099$	$0\cdot095$
$1\frac{1}{2}$	$0\cdot08$	$-0\cdot490$	$-0\cdot068$	$0\cdot108$
$\frac{1}{2}$	$0\cdot04$	$-0\cdot497$	$-0\cdot034$	$0\cdot115$
$-\frac{1}{2}$	0	$-0\cdot500$	0	$0\cdot118$
$-1\frac{1}{2}$	$-0\cdot04$	$-0\cdot497$	$0\cdot034$	$0\cdot115$

TABLE 7 (*Continued*)

Effect of Parameters x_4, x_5

(a) $x_2 = x_3 = 0$	μ'_{22}			μ'_{33}		
x_4 → / x_5 ↓	0	1·0	2·4	0	1·0	2·4
−1·4	1·02	—	—	0·81	—	—
0	1·00	1·03	—	1·00	1·05	—
1	1·01	1·05	—	1·24	1·25	—
4	1·16	1·20	1·40	2·50	2·39	2·15
6·4	1·42	—	—	4·27	—	—

(b) $x_3 \neq 0$	x_3	x_5	μ'_{22}	μ'_{33}
	−1	+6	1·07	2·91
	+1	+6	1·69	4·38
	+1·3	−3	0·88	0·45

(c) $x_2 \neq 0$	x_2	x_4	μ'_{22}	μ'_{33}
	−1·2	+2·2	0·76	0·84

(d) Higher terms included	$\xi_3 = \varepsilon$		μ'_{22}	μ'_{33}
	$x_5 = +7\cdot7$ ($+$term in $\tau_9(\varepsilon)$)		1·71	7·05
	$x_4 = +2\cdot3$ ($+$term in $\tau_6(\varepsilon)$)		1·14	0·99

(c) The effect of the parameter x_2 is to allow slightly higher values of x_4 when x_5 is small (maximum $x_4 = +2\cdot2$ when $x_2 = -1\cdot2$).

(d) Somewhat higher values of x_4, x_5 are possible if higher order terms are added to equation (44). The maximum value of x_5 is unaffected by the addition of a term $\tau_7(\varepsilon)$, and is increased to $+7\cdot7$ by the addition of a term $\tau_9(\varepsilon)$. The maximum value of $|x_4|$ is not greatly altered by the addition of terms $\tau_6(\varepsilon)$ and $\tau_8(\varepsilon)$.

The quantities μ'_{jj}, obtained by calculating the values $\mu_{jj} = \Sigma T_j^2(x)/\Sigma T_j^2(\varepsilon)$ and dividing by the tabulated values μ_{jj} for the same x_2, x_3, provide a measure of the deviations of the standard errors from the exact least squares values. Values of μ'_{jj} for various values of the parameters are listed in Table 7.

The general conclusions to be drawn from this table concerning the probable deviations of the standard errors from the exact least squares values might be set out as follows :

		2nd Degree Polynomial	3rd Degree Polynomial
(i) Likely deviations	..	3%	12%
(ii) Possible deviations	..	15%	50%
(iii) Improbable deviations	·	30%	Factor of 2

It should be realized that high values of x_5 correspond to a very wide spacing at the ends of the range and a very close spacing near $\varepsilon = \pm\frac{1}{4}(n-1)$, and so very high values of x_5 are not likely to occur in practical examples.

<div align="center">TABLE 8</div>

<div align="center">VALUES OF μ_{jj} FOR SMALL n</div>

(a) $x_3 = +1$	n	9	16	25	∞
	μ_{11}	0·652	0·656	0·657	0·657
	μ_{22}	0·714	0·761	0·772	0·781
	μ_{33}	0·537	0·675	0·713	0·745

(b) $x_3 = -1$	n	9	16·	25	∞
	μ_{11}	1·458	1·457	1·457	1·457
	μ_{22}	1·444	1·408	1·401	1·394
	μ_{33}	1·564	1·414	1·386	1·361

The parameters x_4, x_5 will also produce errors in the rough approximations to φ, x_2, x_3, obtained by the use of equations (11) to (13). The approximation to x_2 is found to be increased by $0\cdot10x_4$, and the approximation to φ to be reduced by $0\cdot05x_5$. The approximation to x_3 is increased by an amount varying from $0\cdot05x_5$ to $0\cdot15x_5$, the larger amount corresponding to the cases in which x_3 is negative.

When x_4, x_5 are small, the effect they have on μ_{22}, μ_{33} can be estimated from the leading terms in these parameters. These terms are found to be :

<div align="center">(i) for μ_{22}, $0\cdot0496x_4{}^2$;</div>

<div align="center">(ii) for μ_{33}, $0\cdot156x_4{}^2 + 0\cdot192x_5$.</div>

For the examples discussed in Section VI, the parameters and the corrections to μ_{33} estimated from the leading terms are set out below.

(i) Example 1

$$x_4 = -0\cdot522, \quad x_5 = +1\cdot802.$$
$$\mu_{33} = 1\cdot204 + 0\cdot390 = 1\cdot59 \text{ (exact } 1\cdot58).$$

(ii) Example 2

$$\varkappa_4 = +0 \cdot 061, \quad \varkappa_5 = -0 \cdot 094.$$
$$\mu_{33} = 0 \cdot 820 - 0 \cdot 017 = 0 \cdot 80 \quad (\text{exact } 0 \cdot 77).$$

(iii) Example 3

$$\varkappa_4 = +0 \cdot 107, \quad \varkappa_5 = +0 \cdot 606.$$
$$\mu_{33} = 1 \cdot 130 + 0 \cdot 119 = 1 \cdot 25 \quad (\text{exact } 1 \cdot 23).$$

Only in Example 1. where the parameter $\varkappa_5$ is rather large, is the correction to μ_{33} important.

The effect produced by the neglect of terms j^2/n^2 in the calculation of the quantities μ_{jj} also requires consideration. Evaluation of the true values μ_{jj} for small values of n leads to the figures shown in Table 8. From this table it is seen that the dependence of μ_{jj} on n is small, except for the quantity μ_{33} when n is less than 16.

VIII. Conclusion

There are two separate questions which require some discussion. Firstly, whether the standard errors obtained from the exactly calculated values φ, $\varkappa_2$, $\varkappa_3$ are in good agreement with the true least squares values. Secondly, whether the standard errors obtained by rough estimation of the parameters are of use in practical examples.

As regards the first point, it seems clear that the standard errors of a_1, a_2 will differ from the calculated least squares values by at most a few per cent. In the examples given in Section VI the difference is less than 1 per cent. in each case. For the coefficient a_3 differences of up to 10 per cent. will commonly occur, and larger differences may occur in certain cases. In the examples given here the differences are 13, 3, and 4 per cent. For the fitted values when the polynomial is of the second degree the agreement should always be to within a few per cent.—in the examples given the differences are always less than 2 per cent. When the polynomial is of the third degree, the percentage differences will be less than the difference for the coefficient a_3 in the region of interpolation ($|k| < 1$), and of the same order as for this coefficient in the region of extrapolation ($|k| > 1$). For the examples of Section VI the maximum deviations are 11, 3, and 2 per cent.

As was shown in Section VII, the principal cause of these divergences is the neglect of the higher order parameters $\varkappa_4$ and $\varkappa_5$. It does not seem feasible, however, to include the effects of these parameters in a treatment of the kind given here. The necessity for calculating four parameters and for tabulating the various quantities in terms of them would make the treatment so time-consuming and involved that it would be of little practical use. Consequently it is felt that the present treatment will provide a picture of the behaviour of the standard errors of quantities occurring in curve fitting by the method of least squares which is sufficiently detailed to serve as a basis for the calculations of the efficiencies of other methods of fitting.

Concerning the second point mentioned in the opening paragraph of this section, the principal use of the methods described here would be in the estimation of the standard errors of the fitted values. The exact calculation of these errors for a number of points is somewhat tiresome, since the orthogonal polynomials have to be determined for each point. For this reason Table 6 should prove of use in practical examples where estimates of the standard errors of the fitted values are required. It should be remembered, however, that the estimates cannot be relied on when a third degree polynomial is fitted to a series of observations which are very widely spaced at the extremes of the range and very closely spaced near $\varepsilon = \pm \frac{1}{4}(n-1)$.

For the examples given in Section VI the differences from the exact values are less than 10 per cent. for $|k| < 0.8$, and less than 15 per cent. for $|k| > 0.8$. The larger deviations for $|k| > 0.8$ are due principally to the error in the estimated value of φ.

IX. References

GUEST, P. G. (1950).—*Aust. J. Sci. Res.* A **3** : 173.

GUEST, P. G. (1951).—*Ann. Math. Statist.* **22** : 537.

GUEST, P. G. (1952).—*Aust. J. Sci. Res.* A **5** : 238.

JAEGER, W., and VON STEINWEHR, H. (1921).—*Ann. Phys. Lpz.* **64** : 305.

KENDALL, M. G. (1948).—" The Advanced Theory of Statistics." 2nd Ed. Vol. 2. p. 150. (C. Griffin and Co. : London.)

CALCULATION OF ACCURACY OF RESULTS OF GRAPHICAL SQUARE INTERCOMPARISONS

By P. M. GILET* and G. S. WATSON†

Summary

Graphical squares of two types are used for calculating the results of inter-comparisons of standards of length, mass, and other quantities, and the results of calibration of scales. By means of graphical squares, estimates of the true values or " improved " values are obtained from observations. From the differences between the improved and observed values, the residuals, a measure of the accuracy of the intercomparison, may be obtained. In this paper expressions are derived which enable the calculation, from the residuals, of accuracies of observed values, of improved values, and of various quantities derived from the improved values.

I. INTRODUCTION‡

In the maintenance of standards, much use is made of intercomparisons by which several nominally equal standards are compared in all possible pairs. The results of such comparisons in pairs are used to form a so-called graphical square. The graphical square enables the calculation by simple arithmetic of improved values for the differences between the standards. Such a graphical square is illustrated by Johnson (1923, Fig. 19).

From the residuals obtained by subtracting the observed value for each difference between a pair of standards from the corresponding improved value, it should be possible, of course, to calculate the accuracy both of the observed values and of the improved values. Johnson, however, gives no indication of any method whereby the accuracies can be calculated from the residuals.

One of the uses of the type of square mentioned above is the intercomparison of the main divisions of a standard scale. However, this type of square is not suitable for the calibration of the subdivisions of the main divisions of a scale,

* Division of Metrology. C.S.I.R.O.: present address: Department of Mechanical Engineering, University of Melbourne.

† Department of Statistics, University of Melbourne.

and for this work a second type of graphical square is used, resulting from the comparison of each subdivision of one main division with each subdivision of another main division. Such a square of the second type is shown in Figures 20 and 21 of the above-mentioned article (Johnson 1923).* But here again no mention is made of any method for calculating accuracies from the residuals.

The discussion in this paper will be confined mainly to the calibration of scales, circular and linear, but the expressions derived are applicable to inter-comparisons of quantities of many kinds. For example, although one set of expressions has been derived by consideration of the calibration of the main divisions of a linear scale, the set of expressions is equally applicable to the results of comparing, say, four 3 in. slip gauges in combination against a known 12 in. end bar, and of intercomparing the four slip gauges in pairs, or of comparing, say, five 200 g. masses against a known kilogram and of intercomparing the five 200 g. masses in pairs.

II. Accuracy of Results from a Square of the First Type
(a) Procedure for Use of Square

The method of using a graphical square of the first type is given in detail by Johnson (1923) and for convenience is summarized below. The method is described in the above article for the case of 10 quantities, but in the following the general case is taken where there are n quantities intercompared in all possible pairs. Table 1 represents a partly completed square for n quantities. $\alpha_1, \alpha_2, \ldots, \alpha_n$ are the true values of the quantities or, since we are usually comparing nominally equal quantities, the true departures from their nominal values. For definiteness, the first interpretation will be used below. $e_{12}, e_{13}, \ldots, e_{n-1\,n}$ are the errors in measurement made in the comparisons.

The observed differences between quantities, such as $y_{12} = \alpha_1 - \alpha_2 + e_{12}$ are entered, as shown, below the diagonal line, and again entered with a change of sign in the appropriate spaces above the diagonal.

The observed differences are added in columns and the sums entered in the M row. Each sum is divided by n and the resulting mean values are entered in the $\bar{M}$ row.

The square is now at the stage shown in Table 1, excluding the last two rows.

The improved value for a difference between any pair of quantities is given by the difference between the corresponding $\bar{M}$ values. It can be shown that, in fact, the graphical square gives a least squares solution for the $\frac{1}{2}n(n-1)$ observational equations. For example, $(\bar{M}_1 - \bar{M}_2)$ is the best (least squares) value corresponding to the true difference $(\alpha_1 - \alpha_2)$.

The values of these improved estimates, $(\bar{M}_1 - \bar{M}_2)$, $(\bar{M}_1 - \bar{M}_3)$, etc., are calculated and entered in the appropriate spaces of the square. Each improved estimate is subtracted from the corresponding observed difference and the resulting residual also entered in the appropriate space.

---

* An example of a square of this type is given in Table 3 of this paper.

These residuals, as stated by Johnson, are a measure of the accuracy of the whole comparison. In the following, expressions are derived enabling accuracies to be calculated from the residuals.

TABLE 1

SQUARE OF THE FIRST TYPE

	α_1	α_2		α_{n-1}	α_n
α_1		$-y_{12}$		$-y_{1\ n-1}$	$-y_{1\ n}$
α_2	y_{12}			$-y_{2\ n-1}$	$-y_{2\ n}$
α_{n-1}	$y_{1\ n-1}$	$y_{2\ n-1}$			$-y_{n-1\ n}$
α_n	$y_{1\ n}$	$y_{2\ n}$		$y_{n-1\ n}$	
M	M_1	M_2		M_{n-1}	M_n
$\overline{M}$	$\overline{M}_1$	$\overline{M}_2$		$\overline{M}_{n-1}$	$\overline{M}_n$
M adjusted for departure of overall length	$\overline{M}_1+\dfrac{A}{n}$	$\overline{M}_2+\dfrac{A}{n}$		$\overline{M}_{n-1}+\dfrac{A}{n}$	$\overline{M}_n+\dfrac{A}{n}$
Progressive total ..	$P_1=$ $\overline{M}_1+\dfrac{A}{n}$	$P_2=\overline{M}_1$ $+\overline{M}_2+2\dfrac{A}{n}$		$P_{n-1}=\overline{M}_1+\dots$ $+\overline{M}_{n-1}+(n-1)\dfrac{A}{n}$	$P_n=\overline{M}_1+\dots$ $+\overline{M}_n+A=A$

Reverting to the procedure for obtaining the results and leaving for the moment the matter of their accuracy, it will be noted that the procedure has been outlined to the stage of obtaining improved estimates of the differences between the quantities concerned. In many applications this is all that is necessary. For example, when several end bars, several masses, or several overall lengths of line standards have been intercompared, the aim is simply to obtain improved estimates of the differences between the quantities. One

or more of the quantities is known, and the values for the others can be deter-
mined from the known values and the appropriate improved estimates of the
differences.

It happens sometimes, however, that the sum of the quantities compared
is known, but none of the individual quantities.   This would happen, for example,
when five unknown 200 g. masses have been compared with a known kilogram
and then intercompared in pairs.   The sum of the five 200 g. masses would then
be known, as would the improved estimates of their differences.   Again, in the
calibration of the main divisions of a line standard, the divisions are inter-
compared in pairs and improved estimates obtained for their differences.   The
overall length of the standard is known, and it is necessary to determine the
value of each main division from the known sum of the main divisions (i.e. the
overall length) and the improved estimates of the differences.

In such cases the square is extended as indicated in the last two rows of
Table 1 where A is the value assumed for the sum of the quantities $\alpha_1, \alpha_2, \ldots, \alpha_n$.

In the example given above relating to five 200 g. masses, the improved
value for each mass would be given by the appropriate space in the second last
row of Table 1.

In the case of the line standard, the improved value for each main division
would be given by the second last row, and the last row would give improved
values for all the main intervals of the standard starting from one end.

If the procedure given by Johnson for the intercomparison of the main
divisions of a line standard is followed exactly, it will be found that compartments
of the square which are diagonally adjacent contain a common error and are not
independent.   If all the observations are repeated, it is possible to choose
observations alternately from the two sets and combine them to form two squares
in each of which the observed values are now wholly independent.   The results
from the two squares are used as a check upon each other.

The analysis by the methods set out below applies to either of the two
squares formed as described.

(b) Statistical Investigation of a Square of the First Type

Denoting the observed differences between quantities whose true values
are α_i and α_j by y_{ij}, it is seen from Table 1 that

$$y_{ij} = \alpha_i - \alpha_j + e_{ij}, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad \textbf{(2.1)}$$

with $i = 1, 2, \ldots, n-1$ and $j = i+1, i+2, \ldots, n$.   Let it be supposed that all
the errors e_{ij} are independent random variables with zero means and a common
variance σ_0^2.

If the true value of the sum of all the quantities is known, A say, then the
α's must be estimated from the observations subject to the restriction that
$\sum_1^n \alpha_i = A$.   The best linear unbiased estimates may be obtained by the method
of least squares.*

---

* All the statistical concepts and theorems used below may be found, *inter alia*, in Mood
(1950).

The normal equations are found by minimizing

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (y_{ij} - x_i + x_j)^2$$

by choice of $\alpha_1, \alpha_2, \ldots, \alpha_n$ with $\sum_1^n \alpha_i = A$. They are easily seen to be

$$M_j = n\hat{\alpha}_j - \sum_1^n \hat{\alpha}_j \qquad (j=1, \ldots, n), \qquad \ldots \ldots \ldots (2.2)$$

that is,

$$\hat{\alpha}_j = \bar{M}_j + \frac{A}{n}. \qquad \ldots \ldots \ldots \ldots \ldots \ldots \ldots (2.3)$$

The residuals v_{ij} are defined by

$$v_{ij} = y_{ij} - \hat{\alpha}_i + \hat{\alpha}_j = y_{ij} - M_i + M_j, \qquad \ldots \ldots \ldots \ldots (2.4)$$

and an estimate S_0^2 of σ_0^2 can be found by dividing the sum of squares of these residuals by the number of degrees of freedom associated with it. This number is given by the general rule :

degrees of freedom = (number of observations)
  — (number of independent constants estimated).

In the present instance, there are

$$\frac{n(n-1)}{2} - (n-1) = \frac{(n-1)(n-2)}{2}$$

degrees of freedom. Thus the estimate of σ_0^2 is

$$S_0^2 = 2 \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} v_{ij}^2}{(n-1)(n-2)}. \qquad \ldots \ldots \ldots \ldots (2.5)$$

By using the equation (2.4) it may be verified that

$$\sum_{i=1}^{n-1} \sum_{=i+1}^{n} v_{ij}^2 = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} y_{ij}^2 - \frac{1}{n} \sum_{j=1}^{n} M_j^2. \qquad \ldots \ldots (2.6)$$

If a calculating machine is being used, (2.6) leads to the most expeditious method for computing S_0^2.

It is now necessary to find the variances of the estimates of α's and functions of them. In their derivations, account must be taken of the fact that every pair of M_i's has a term in common.
Thus

$$\text{var } (\hat{\alpha}_i - \hat{\alpha}_j) = \frac{2\sigma_0^2}{n} \qquad \begin{matrix} (i=1, \ldots, n-1) \\ (j=1+i, \ldots, n) \end{matrix} \Big\}, \quad \cdot (2.7)$$

$$\text{var } (P_r) = \text{var } (\hat{\alpha}_1 + \ldots + \hat{\alpha}_r)$$

$$= \frac{r(n-r)}{n^2} \sigma_0^2 \qquad (r=1, \ldots, n). \qquad \ldots \ldots (2.8)$$

In particular, (2.8) with $r=1$ gives

$$\text{var } (\hat{\alpha}_i) = \frac{n-1}{n} \sigma_0^2 \qquad (i=1, \ldots, n). \qquad \ldots \ldots (2.9)$$

C

The case of a circular scale may be given as an example in which the sum of the unknown quantities is known exactly. Here the unknown quantities are the true values of the main divisions of the $360°$ used for a square of the first type.

For a linear scale or a group of equal masses the sum of the unknown quantities is never known exactly but merely through another measurement. Then A will be subject to error with variance $\sigma_A{}^2$, say. Under these circumstances, the best method of estimation is complicated. The most practical method is to use the above estimates of the α's (2.3) and take account of the error in A when finding the variance formulae. Then the procedure and results are exactly as before except (2.8). This becomes now

$$\text{var } (P_r) = \frac{1}{n^2}[r(n-r)\sigma_0{}^2 + r^2\sigma_A{}^2] \qquad (r=1, \ldots, n). \ \ldots (2.10)$$

Again, in particular, (2.10) gives with $r=1$

$$\text{var } (\hat{\alpha}_i) = \frac{1}{n^2}[(n-1)\sigma_0{}^2 + \sigma_A{}^2] \qquad (i=1, \ldots, n). \ \ldots (2.11)$$

To use (2.10) or (2.11) an estimate of $\sigma_A{}^2$ must be available from a previous investigation.

Naturally $S_0{}^2$ of (2.5) and (2.6) will be used for $\sigma_0{}^2$ in (2.7) to (2.11).

Thus, for a linear scale, the best estimates of the main divisions are found from a square as in Table 1 and the estimates of the accuracies of the quantities of interest are found from (2.5), (2.6), (2.7), (2.10), and (2.11).

For a circular scale the same square is used but (2.8) and (2.9) replace (2.10) and (2.11) in accuracy calculations. It will be noted that

$$\text{var } (P_{n-r}) = \text{var } (P_r)$$

and that the maximum value of var (P_r) occurs for $r=n/2$ (n even) and $r=(n\pm1)/2$ (n odd). This is illustrated by the broken line curve of Figure 2.

When several nominally equal standards are compared, the last row of the square in Table 1 is omitted since it is devoid of interest.

III. ACCURACY OF RESULTS FROM A SQUARE OF THE SECOND TYPE
(a) Description of Square

This type of square is used when each subdivision of a main division of a scale has been compared with each subdivision of another main division.

Such a square is shown in Figures 20 and 21 of the previously mentioned article (Johnson 1923), and the method is described for the case of two decimetre divisions of a metre scale, each subdivided into 10 centimetre divisions.* In the following the general case is taken of a scale divided and numbered as in Figure 1.

The graphical square of Table 2 refers to the comparison of each of the m subdivisions of the $(P+1)$th main division with each of the m subdivisions of

---

* In Table 3 is given another example of a square of this type.

the $(Q+1)$th main division. The method of representation of divisions is such that, for example, the $(P+1)$th main division is represented by $P/P+m$, and, for example, the third subdivision of the $(Q+1)$th main division is represented by $Q+2/Q+3$.

<div align="center">TABLE 2</div>

<div align="center">SQUARE OF THE SECOND TYPE</div>

	$P/P+1$	$P+1/P+2$		$P+m-1/P+m$	S	M	$-M+\dfrac{D_Q}{m}+\dfrac{G}{m^2}$
$Q/Q+1$	z_{11}	z_{21}		z_{m1}	R_1	$\bar{R}_1$	$\dfrac{G}{m^2}+\dfrac{D_Q}{m}-\bar{R}_1$
$Q+1/Q+2$	z_{12}	z_{22}		z_{m2}	R_2	$\bar{R}_2$	$\dfrac{G}{m^2}+\dfrac{D_Q}{m}-\bar{R}_2$
$Q+m-1/Q+m$	z_{1m}	z_{2m}		z_{mm}	R_m	$\bar{R}_m$	$\dfrac{G}{m^2}+\dfrac{D_Q}{m}-\bar{R}_m$
S	C_1	C_2		C_m	G		
M	$\bar{C}_1$	$\bar{C}_2$		$\bar{C}_m$	$\dfrac{G}{m}$		
$M+\dfrac{D_P}{m}-\dfrac{G}{m^2}$	$\bar{C}_1+\dfrac{D_P}{m}-\dfrac{G}{m^2}$			$\bar{C}_m-\dfrac{D_P}{m}-\dfrac{G}{m^2}$			
Progressive totals ..	P_1	P_2		P_m			

In connection with Table 2 the true lengths of the subdivisions of $P/P+m$ are represented by $\delta_1, \delta_2, \ldots, \delta_m$ and the true lengths of the subdivisions of $Q/Q+m$ are represented by $\theta_1, \theta_2, \ldots, \theta_m$. A typical entry in the table, z_{ij}, stands for the observed difference between the ith subdivision of $P/P+m$ and the jth subdivision of $Q/Q+m$. If the error in the comparison is e_{ij} then
$$z_{ij}=\delta_i-\theta_j+e_{ij}.$$

The rows and columns marked S and M contain respectively the sums and means of observed differences, the grand total being shown as G. The second last row gives the mean values $\bar{C}_j$ adjusted to give the improved values of the lengths of the subdivisions of $P/P+m$ as derived below. It will be noted that the adjustment consists of the addition of D_P/m (where D_P is the estimate of the main division $P/P+m$ obtained from a square of the first type) and of the subtraction of G/m^2. G/m, being equal to $(\Sigma_i\hat{\delta}-\Sigma_j\hat{\theta})$, is the difference between the main divisions $P/P+m$ and $Q/Q+m$ as obtained from the square of Table 2. It can be shown that the adjustment for obtaining the second last row of Table 2 is identical with that given by Johnson (1923, Fig. 20).

Fig. 1

The last column of Table 2 gives the mean values $\bar{R}_j$ adjusted to give the improved values of the subdivisions of $Q/Q+m$ as derived below. It can again be shown that the adjustment is identical with that given by Johnson. The last row of Table 2 gives progressive totals of the row above, i.e. the improved values for the intervals $P/P+1$, $P/P+2$, . . ., $P/P+m$ of the scale. The corresponding column has been omitted from Table 2.

If the procedure given by Johnson for obtaining the observed differences is followed exactly, the errors e_{11}, e_{12}, . . . will not be independent. As described in Section II, two intercomparisons can be made and alternate entries from the intercomparisons made into two squares such as Table 2. If this is done the errors e_{11}, e_{12}, . . . in each square are independent, the results from one of the squares giving a valuable check on the results of the other. The analysis given below is based on the assumption that the errors are independent.

(b) Statistical Investigation of Square of Second Type

Denoting the observed differences of the subdivisions by z_{ij}, the observations in Table 2 may be written as

$$z_{ij}=\delta_i-\theta_j+e_{ij}, \qquad (i,j=1, 2, . . ., m) \ \ . . . . \ \textbf{(3.1)}$$

where the e_{ij}'s are assumed to be independent with zero means and a common variance σ_s^2. As in Section II, the most satisfactory method of estimation is least squares, modified to take account of the estimates of the lengths of the main divisions.

$\overset{m}{\underset{1}{\Sigma}}\delta_i$ and $\overset{m}{\underset{1}{\Sigma}}\theta_j$ represent the true lengths of the main divisions which will have been estimated from a square of the first type. Let the appropriate estimates be

$$D_P=\hat{\alpha}_{r+1}, \quad D_Q=\hat{\alpha}_t \qquad (t\neq r+1)$$

in the notation of Section II (b). Thus, $\delta_1, \delta_2, \ldots, \delta_m$ and $\theta_1, \theta_2, \ldots, \theta_m$ would ideally be estimated by least squares subject to the restrictions

$$\sum_1^m \delta_i = \hat{\alpha}_{r+1}, \quad \sum_1^m \theta_j = \hat{\alpha}_t.$$

The normal equations, obtained as before, are

$$\left. \begin{array}{ll} R_j = \sum_1^m \hat{\delta}_i - m\hat{\theta}_j & (j=1, 2, \ldots, m), \\[2mm] C_i = m\hat{\delta}_i - \sum_1^m \hat{\theta}_j & (i=1, 2, \ldots, m). \end{array} \right\} \quad \cdots (3.2)$$

It is evident that only one restriction can be imposed upon the solutions of (3.2). Let it be that

$$\sum_1^m \hat{\delta}_i = \hat{\alpha}_{r+1}. \quad \cdots\cdots\cdots\cdots\cdots\cdots \quad (3.3)$$

Then

$$\hat{\theta}_j = \frac{\hat{\alpha}_{r+1}}{m} - \bar{R}_j, \quad \cdots\cdots\cdots\cdots\cdots \quad (3.4)$$

$$\hat{\delta}_i = \bar{C}_i + \frac{\hat{\alpha}_{r+1}}{m} - \frac{G}{m^2}. \quad \cdots\cdots\cdots\cdots\cdots \quad (3.4')$$

If the restriction

$$\sum_1^m \hat{\theta}_j = \hat{\alpha}_t \quad \cdots\cdots\cdots\cdots\cdots\cdots \quad (3.5)$$

is used, the estimates are

$$\hat{\delta}_i = \bar{C}_i + \frac{\hat{\alpha}_t}{m}, \quad \cdots\cdots\cdots\cdots\cdots \quad (3.6)$$

$$\hat{\theta}_j = \frac{G}{m^2} + \frac{\hat{\alpha}_t}{m} - \bar{R}_j. \quad \cdots\cdots\cdots\cdots \quad (3.6')$$

The estimates used by Johnson are (3.4') and (3.6').

The residuals are given by

$$w_{ij} = z_{ij} - \hat{\delta}_i + \hat{\theta}_j,$$

where $\hat{\delta}_i$ and $\hat{\theta}_j$ are defined by (3.4) and (3.4') or (3.6) and (3.6'). Thus

$$w_{ij} = z_{ij} - \frac{R_i}{m} - \frac{C_j}{m} + \frac{G}{m^2}. \quad \cdots\cdots\cdots\cdots \quad (3.7)$$

These are the residuals obtained by the method of calculation given by Johnson. It may be shown that

$$\sum_{i=1}^m \sum_{j=1}^m w_{ij}^2 = \sum_{i=1}^m \sum_{j=1}^m z_{ij}^2 - \frac{\Sigma R_i^2}{m} - \frac{\Sigma C_j^2}{m} + \frac{G^2}{m^2}. \quad \cdots\cdots \quad (3.8)$$

This sum of squares carries

$$m^2 - (2m-1) = (m-1)^2$$

degrees of freedom, so that the estimate of σ_s^2, S_s^2, is given by

$$S_s^2 = \frac{\displaystyle\sum_{i=1}^m \sum_{j=1}^m w_{ij}^2}{(m-1)^2}. \quad \cdots\cdots\cdots\cdots\cdots \quad (3.9)$$

When computing (**3.9**) with a calculating machine, use should always be made of (**3.8**).

The variances of the estimates $\hat{\theta}_j$, $\hat{\delta}_i$ from (**3.4**) and (**3.4′**) are found to be

$$\text{var}\ (\hat{\theta}_j) = \frac{\text{var}\ (\hat{\alpha}_{r+1})}{m^2} + \frac{\sigma_s{}^2}{m}, \quad \cdots\cdots\cdots \quad (\mathbf{3.10})$$

$$\text{var}\ (\hat{\delta}_i) = \frac{\text{var}\ (\hat{\alpha}_{r+1})}{m^2} + \frac{(m-1)}{m^2}\sigma_s{}^2. \quad \cdots\cdots \quad (\mathbf{3.11})$$

When the estimates from (**3.6**) and (**3.6′**) are used, the terms containing $\sigma_s{}^2$ in (**3.10**) and (**3.11**) are interchanged. When the procedure described by Johnson is followed, both estimates have the variance

$$\frac{\text{var}\ (\hat{\alpha}_{r+1})}{m^2} + \frac{(m-1)}{m^2}\sigma_s{}^2.$$

The variance of the estimate $\hat{\delta}_1 + \hat{\delta}_2 + \ldots + \hat{\delta}_z = P_z$ of the sum of the lengths of the first z subdivisions is given by

$$\text{var}\ (P_z) = \frac{1}{m^2}[z(m-z)\sigma_s{}^2 + z^2\ \text{var}\ (\hat{\alpha}_{r+1})]. \quad \cdots\cdots \quad (\mathbf{3.12})$$

IV. Accuracy of the Estimated Values of Intervals of a Scale, Each Interval Starting from Zero Graduation

The full procedure for the calibration of a scale (linear or circular) can now be summarized.

The scale is regarded as being composed of n main divisions $\alpha_1, \alpha_2, \ldots, \alpha_n$, each of which has m subdivisions.

The main divisions are intercompared and the results worked up in a square of the first type. This square gives estimates of $\alpha_1, \alpha_2, \alpha_3, \ldots$ and also of $\alpha_1, \alpha_1 + \alpha_2, \alpha_1 + \alpha_2 + \alpha_3, \ldots$

Two divisions are chosen (the choice being made on non-statistical grounds) and the m subdivisions of one are compared with the m subdivisions of the other. These divisions are denoted by $P/P+m$ and $Q/Q+m$ so that the subdivisions are denoted by $P/P+1$, $P+1/P+2$, $\ldots$, $P+m-1/P+m$ and $Q/Q+1$, $Q+1/Q+2$, $\ldots$, $Q+m-1/Q+m$. In Section III, the true values of these subdivisions were denoted respectively by $\delta_1, \delta_2, \ldots, \delta_m$ and $\theta_1, \theta_2, \ldots, \theta_m$. Estimates of these subdivisions and their progressive totals were obtained in Section III from a square of the second type.

It remains to obtain estimates of the intervals $O/P+1, O/P+2, \ldots, O/P+m$ and $O/Q+1, O/Q+2, \ldots, O/Q+m$, and expressions for accuracy of these estimates. Consider generally the estimation of $O/P+z$. If $P/P+m$ is the $(r+1)$th main division, then

$$O/P+z = (\alpha_1 + \alpha_2 + \ldots + \alpha_r) + (\delta_1 + \delta_2 + \ldots + \delta_z). \quad \cdots \quad (\mathbf{4.1})$$

So that, by (2.3) and (3.4′), estimate of

$$O/P + z = (\hat{\alpha}_1 + \hat{\alpha}_2 + \ldots + \hat{\alpha}_r) + (\hat{\delta}_1 + \hat{\delta}_2 + \ldots + \hat{\delta}_z)$$

$$= \left(\overline{M}_1 + \overline{M}_2 + \ldots + \overline{M}_r + r\frac{A}{n} \right) + \left(\overline{C}_1 + \overline{C}_2 + \ldots + \overline{C}_z + \frac{z\hat{\alpha}_{r+1}}{m} - \frac{zG}{m^2} \right).$$

$$\ldots \ldots \ldots \ldots \ldots \ldots \ldots \text{ (4.2)}$$

Thus the estimate of $O/P + z$ is simply the sum of the two appropriate progressive totals in the squares of Tables 1 and 2.

Now

$$\text{var (est. of } O/P + z) = \text{var } (\hat{\alpha}_1 + \hat{\alpha}_2 + \ldots + \hat{\alpha}_r) + \text{var } (\hat{\delta}_1 + \hat{\delta}_2 + \ldots + \hat{\delta}_z)$$

$$+ 2 \text{ cov } (\hat{\alpha}_1 + \hat{\alpha}_2 + \ldots + \hat{\alpha}_r, \hat{\delta}_1 + \hat{\delta}_2 + \ldots + \hat{\delta}_z).$$

$$\ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \text{ (4.3)}$$

The first two terms of (4.3) are given by (2.10) and (3.12).

To find the last term, it will be noted that

$$\text{cov } (\hat{\alpha}_1 + \ldots + \hat{\alpha}_r, \hat{\delta}_1 + \ldots + \hat{\delta}_z) = \frac{z}{m} \text{ cov } (\hat{\alpha}_1 + \ldots + \hat{\alpha}_r, \hat{\alpha}_{r+1})$$

$$= \frac{rz}{m} \text{ cov } (\hat{\alpha}_1, \hat{\alpha}_{r+1})$$

$$= \frac{rz}{mn^2} (\sigma_A{}^2 - \sigma_0{}^2).$$

Combining results,

$$\text{var (est. of } O/P + z) = \frac{1}{n^2} [r(n-r)\sigma_0{}^2 + r^2\sigma_A{}^2] + \frac{1}{m^2} [z(m-z)\sigma_s{}^2 + \frac{z^2}{n^2} \{(n-1)\sigma_0{}^2 + \sigma_A{}^2\}]$$

$$+ \frac{2rz}{mn^2} [\sigma_A{}^2 - \sigma_0{}^2], \ldots \ldots \ldots \ldots \ldots \ldots \text{ (4.4)}$$

which may be rearranged to give

$$\text{var (est. of } O/P + z) = \frac{1}{m^2 n^2} [\sigma_0{}^2 \{(rm^2 + z^2)(n-r-1) + r(m-z)^2\}$$

$$+ \sigma_s{}^2 n^2 z(m-z) + \sigma_A{}^2 (rm+z)^2]. \ldots \ldots \text{ (4.5)}$$

In the case of a circular scale where A is known exactly, the same result with $\sigma_A{}^2 = 0$ applies.

V. Summary of Symbols and Expressions relating to Accuracy for Graphical Squares of Both Types and for Combined Results from Squares of Both Types

(a) Square of First Type

Number of quantities intercompared $= n$.

Number of comparisons $= \frac{1}{2} n(n-1).$*

Residual, i.e. observed difference minus improved estimate $= v.$

Estimated variance of value used for sum of quantities intercompared $= S_A{}^2.$

* Note that each of the comparisons must be statistically independent of the remainder.

Estimated variance for observed differences $=S_0{}^2=\dfrac{2\Sigma v^2}{(n-1)(n-2)}.$*

$$\Sigma v^2=\Sigma y^2-\frac{1}{n}\Sigma M^2.$$

Estimated variance for improved estimates
of differences $\qquad =\mathrm{var}\ (\hat\alpha_i-\hat\alpha_j)=\dfrac{2}{n}S_0{}^2.$

Estimated variance for improved estimate
of each quantity $\qquad =\mathrm{var}\ (\alpha_i)=\dfrac{1}{n^2}[S_A{}^2+(n-1)S_0{}^2].$

Estimated variance for progressive total of
improved estimates of quantities up to
and including the rth quantity $\qquad =\mathrm{var}\ (P_r)=\dfrac{1}{n^2}[r^2S_A{}^2+r(n-r)S_0{}^2].$

(b) Square of Second Type

Number of subdivisions of a main division $=m.$

Number of comparisons $\qquad =m^2.$†

Residual, i.e. observed difference minus
improved estimate $\qquad =w.$

Estimated variance for improved estimate
of a division from a square of the first type $=\mathrm{var}\ (\hat\alpha_{r+1})=\dfrac{1}{n^2}[S_A{}^2+(n-1)S_0{}^2].$

Estimated variance for observed differences
between subdivisions $\qquad =S_s{}^2=\dfrac{1}{(m-1)^2}\Sigma w^2.$

$$\Sigma w^2=\Sigma z^2-\frac{1}{m}\Sigma R^2-\frac{1}{m}\Sigma C^2+\frac{G^2}{m^2}.$$

Estimated variance for improved estimates
of differences between subdivisions $\qquad =\mathrm{var}\ (\hat\theta_j-\hat\delta_i)=\dfrac{1}{m^2}(2m-1)S_s{}^2.$

Estimated variance for improved value of
each subdivision $\qquad =\mathrm{var}\ (\theta_j)=\mathrm{var}\ (\delta_i)$

$$=\frac{1}{m^2}\ \mathrm{var}\ (\hat\alpha_{r+1})+\frac{1}{m^2}(m-1)^2S_s{}^2.$$

Estimated variance for progressive total of
subdivisions up to and including the zth
subdivision $\qquad =\mathrm{var}\ (P_z)$

$$=\frac{1}{z^2}[z^2\ \mathrm{var}\ (\hat\alpha_{r+1})+z(m-z)S_s{}^2].$$

(c) Intervals from Zero Graduation

Estimated variance for value of interval,
$O/P+z$, this value being obtained by
adding the values for the first r main
intervals, to give O/P, and the value for
the subinterval $P/P+z$ $\qquad =\mathrm{var}\ (\text{est. of } O/P+z)$

$$=\frac{1}{m^2n^2}[S_0{}^2\{(rm^2+z^2)(n-r-1)+r(m-z)^2\}+S_s{}^2n^2z(m-z)+S_A{}^2(rm+z)^2].$$

---

* Note that the number of v's used in obtaining Σv^2 is $\frac{1}{2}n(n-1)$, i.e. the number is that of the comparisons made and not the number of compartments in the square which is $n(n-1)$.

† Note that each comparison must be statistically independent of the remainder.

VI. Example of Graphical Square Method applied to the Calibration of a Circular Scale

The graphical square method of calibration was used some time ago in the calibration of the scale of the 40 in. circular dividing engine of the Division of Metrology made by the Société Génevoise de Physique.

The first stage was the intercomparison of the twelve 30° divisions of the scale, using a graphical square of the first type for calculating the results. The second stage was the comparison of each 5° subdivision of one 30° division with each 5° subdivision of another 30° division. This was done six times to cover the six pairs of 30° divisions, the results being calculated using 6×6 squares of the second type. The third stage was the comparison of each 1° subdivision of one 5° division with each 1° subdivision of another 5° division. This was done 36 times, the results being calculated using 5×5 squares of the second type.

For this example, only the first and second stages will be considered, and the square of the first type used for the intercomparison of the 30° divisions is not reproduced, the method being quite simple. It is considered advisable to reproduce one of the squares of the second type.

The square shown in Table 3 is one of the squares used for the second stage of the calibration. The complete square is shown, in the form given by Johnson. Comparing Table 3 with Table 2, it will be seen that the additional items shown in Table 3 are improved estimates (at the top of each compartment), residuals (at the bottom of each compartment), a column corresponding to the last row of Table 2, and a row and a column giving values for the intervals from zero and making use of the values obtained for the intervals 0/30 and 0/210 with the square of the first type. If the value of Σw^2 is to be calculated using (3.8), there is no need to include in the square improved estimates of differences or residuals.

The observed differences entered in this square (as for all the other squares used) were obtained alternately from two independent sets of values in order that all the errors of measurement in the observed values should be independent. It was permissible, therefore, to make use of the expressions derived in this paper to calculate the accuracies.

The sum of the squares of the residuals in Table 3 is $2 \cdot 0016$ and therefore $\sqrt{\Sigma w^2} = 1 \cdot 415$ sec.

The standard deviation for the observed differences

$$S_s = \frac{1}{m-1} \sqrt{\Sigma w^2}, \qquad \text{with } m = 6,$$

$$= 0 \cdot 28 \text{ sec.}$$

The estimates, $S_s{}^2$ of the variances for the six squares of this stage were pooled, and the pooled value of S_s found to be $0 \cdot 32$ sec.

The standard deviation of the observed differences for the square of the first type, S_0, calculated from the residuals of this square, was found to be $0 \cdot 33$ sec.

TABLE 3
EXAMPLE OF A SQUARE OF THE SECOND TYPE
Values in seconds of arc

	30/35	35/40	40/45	45/50	50/55	55/60	S	M	Add $-\frac{1}{6}d_{30/60}+0{\cdot}17 = +0{\cdot}17$	Successive Addition and Change of Sign	Add Correction for 0/210 $= -0{\cdot}30$
210/215	−0·27 −0·40 +0·13	−0·17 −0·20 +0·03	−0·41 −0·80 +0·39	−0·31 −0·10 −0·21	−0·12 +0·30 −0·42	−0·52 −0·60 +0·08	−1·80	−0·30	−0·13	(210/215) +0·13	(0/215) −0·17
215/220	+0·23 +0·20 +0·03	+0·33 +0·70 −0·37	+0·09 +0·30 −0·21	+0·19 +0·30 −0·11	+0·38 −0·30 +0·68	−0·02 0·00 −0·02	+1·20	+0·20	+0·37	(210/220) −0·24	(0/220) −0·54
220/225	+0·03 −0·20 +0·23	+0·13 0·00 +0·13	−0·11 −0·10 −0·01	−0·01 +0·30 −0·31	+0·18 +0·30 −0·12	−0·22 −0·30 +0·08	0·00	0·00	+0·17	(210/225) −0·41	(0/225) −0·71
225/230	−0·05 0·00 −0·05	+0·05 +0·30 −0·25	−0·19 −0·10 −0·09	−0·09 −0·30 +0·21	+0·10 +0·40 −0·30	−0·30 −0·40 +0·10	−0·50	−0·08	+0·09	(210/230) −0·50	(0/230) −0·80
230/235	−0·10 −0·20 +0·10	0·00 +0·10 −0·10	−0·24 −0·30 +0·06	−0·14 −0·40 +0·26	+0·05 +0·10 −0·05	−0·35 −0·10 −0·25	−0·80	−0·13	+0·04	(210/235) −0·54	(0/235) −0·84
235/240	−0·32 +0·10 −0·42	−0·22 −0·20 −0·02	−0·46 −0·50 +0·04	−0·36 −0·50 +0·14	−0·17 −0·40 +0·23	−0·57 −0·60 +0·03	−2·10	−0·35	−0·18	(210/240) −0·36	(0/240) −0·66
S	−0·50	+0·10	−1·30	−0·70	+0·40	−2·00	−4·00				
M	−0·08	+0·02	−0·22	−0·12	+0·07	−0·33		−0·66			
Add $\frac{1}{6}d_{210/240}+0{\cdot}17 = 0{\cdot}11$	+0·03	+0·13	−0·11	−0·01	+0·18	−0·22			↗ −0·11		
Successive addition	(30/35) +0·03	(30/40) +0·16	(30/45) +0·05	(30/50) +0·04	(30/55) +0·22	(30/60) 0·00					
Add correction for 0/30 $= -0{\cdot}64$	(0/35) −0·61	(0/40) −0·48	(0/45) −0·59	(0/50) −0·60	(0/55) −0·42	(0/60) −0·64					

The third basic standard deviation, S_4, that of the value used for the total length of the scale, is in this case equal to zero.

From the above basic standard deviations, the accuracy of all the results of the calibration were computed using the expressions summarized in Section V.

It is of some interest to show how the accuracies vary for the different intervals starting at the zero graduation.

From Section V, the standard deviations for the improved values of the intervals $0/5, 0/10, \ldots, 0/355$ are given by the expression for var (est. of $O/P+z$) with $S_4=0$, $n=12$, $m=6$,

and $\quad r=1, 2, \ldots, 12,$

$\quad z=1, 2, \ldots, 6.$
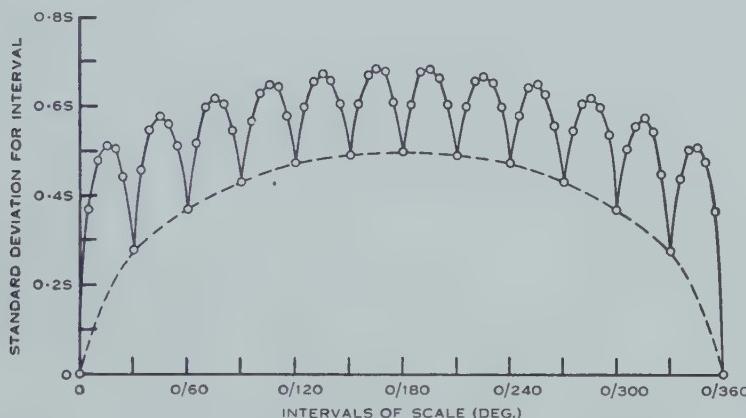
Fig. 2.—Twelve divisions of 30° compared using one graphical square of first type. Twelve subdivisions of 5° compared using each of six graphical squares of second type.

S=standard deviation of observed differences.

It will have been noted that the computed values for S_0 and S_s were closely equal. Assuming, therefore, that S_0 and S_s have a common value S, and substituting for S_4, n, and m, we have

$$S_{0/P+z}=\frac{S}{72}\sqrt{\{432r+864z-12rz-36r^2-133z^2\}}$$

with $r=1, 2, \ldots, 12,$

$\quad z=1, 2, \ldots, 6.$

These values of $S_{0/P+z}$ are shown in the graph of Figure 2. The points joined by the broken curve refer to the intervals obtained solely from the square of the first type.

VII. Acknowledgments

Acknowledgments and thanks are due to the following officers of C.S.I.R.O. Mr. Macintyre of the Section of Mathematical Statistics first indicated the possibility of the work reported by deriving an expression for the accuracy of

improved estimates of differences for a 4×4 square of the first type. Mr. W. A. F. Cuninghame of the Division of Metrology derived the corresponding expression for a generalized $(n \times n)$ square of the first type. Miss P. M. Yelland rendered valuable assistance in checking some of the later work and Miss J. M. Totolos assisted greatly in preparing the paper for publication. Thanks are also due to Dr. Cornish of the Section of Mathematical Statistics for checking for accuracy all the derivations and expressions given in the original form of the paper.

VIII. References

JOHNSON, W. H. (1923).—Comparators ; in " Dictionary of Applied Physics." (Edited by Sir Richard Glazebrook.) Vol. 3. pp. 232-57. (Macmillan & Co. : London.)

MOOD, A. M. (1950).—" Introduction to the Theory of Statistics." (McGraw-Hill : New York.)

ENERGY RESOLUTION OF γ–RADIATION UP TO 18 MeV. BY SODIUM IODIDE SCINTILLATION COUNTERS

By J. G. CAMPBELL* and A. J. F. BOYLE†

Summary

The pulse-height distributions to be expected from scintillation counters using thallium-activated sodium iodide crystals have been investigated for γ-radiation of energies 6, 12, and 18 MeV. with two sizes of cylindrical crystal (2·5 cm. long by 2·5 cm. diameter and 7·5 cm. long by 4 cm. diameter). A Monte Carlo calculation of the partial shower process by which photons are absorbed in the crystals has given the probability distributions of ionization energy from homogeneous incident photons. The spreading of the distributions by statistical effects in the photomultiplier has been measured using light flashes of constant integrated intensity from a rotating mirror. The curves predicted by combining these two factors are compared with experimental pulse-height distributions for γ-radiation from proton bombardment of fluorine, boron, and lithium.

It is concluded that energy resolution above 10 MeV. is severely limited by broadening of peaks introduced by an appreciable and variable fraction of the energy of each incident photon escaping from the crystal as bremsstrahlung. This could be minimized by use of crystals with dimensions large compared with 8 cm., the maximum mean free path of bremsstrahlung photons in sodium iodide.

I. INTRODUCTION

Hofstadter and McIntyre (1950a, 1950b) have shown that for energies of the order of 1 MeV. the pulse-height distributions obtained from a scintillation counter using a thallium-activated sodium iodide crystal can be interpreted in terms of the energy distributions of secondary electrons from the γ-capture processes of photoelectric effect, Compton scattering, and pair production. Maeder and Wintersteiger (1952a, 1952b) have shown the further effect of capture of degraded Compton-scattered quanta in the crystal.

When such a counter is used with γ-radiation above 6 MeV. it is found that the peaks in the pulse-height distributions are considerably broader than can be accounted for by the above-mentioned processes alone. The work described herein was undertaken to ascertain whether the additional broadening could be explained by a more complete investigation of the partial shower process which takes place when each high energy quantum is absorbed.

The use of a scintillation counter as a γ-ray spectrometer depends on the approximate proportionality which exists between the output pulse height and the energy of the incident photons. By breaking down the operation of the

* Physics Department, University of Melbourne.

† Dunlop Research Scholar, University of Melbourne ; present address : Research School of Physical Sciences, Australian National University.

counter into consecutive stages it is possible to see where significant departures from strict proportionality can arise. Considering only homogeneous γ-quanta of energy E, possible sources of variation in output pulse height are :

   (i) The energy I given up by ionization in the crystal. We will let $f(I,E)\mathrm{d}I$ denote the probability that a quantum loses energy between I and $I+\mathrm{d}I$.

  (ii) The number of light photons emitted per unit energy of ionization in the crystal. In sodium iodide this number is independent of I for electrons above 1 keV. (Taylor *et al.* 1951), and any variation in fluorescence efficiency through the crystal will here be neglected. Since, furthermore, the number is large, purely statistical variation is small, and so it will be assumed that the number of photons emitted is proportional to I.

 (iii) The fraction of the light photons which reach the photosensitive surface of the photomultiplier. The main variation in this arises from the dependence of the efficiency of light collection upon the position in the crystal from which the light emanates. This can be due to imperfections in the optical properties of the crystal (Garlick and Wright 1952) or to differences in the effective solid angle subtended by the photosurface. For the purposes of the present paper, this source of variation will also be neglected, and the number of photons reaching the photosensitive surface will be taken to be aI, where a is constant.

  (iv) The fraction of photons reaching the photosensitive surface which eject electrons, and

   (v) The multiplication of electrons in the multiplier.

The last two factors have been considered in detail by various workers (Seitz and Mueller 1950 ; Garlick and Wright 1952). We will let $g(p,aI)\mathrm{d}p$ denote the probability that aI photons reaching the photosensitive surface will give rise to an output pulse between p and $p+\mathrm{d}p$ in height.

The probability that a γ-quantum of energy E absorbed in the crystal gives rise to an output pulse from the photomultiplier between p and $p+\mathrm{d}p$ is now $h(p,E)\mathrm{d}p$, where

$$h(p,E)=\int_{0}^{E} g(p,aI)f(I,E)\mathrm{d}I. \quad \ldots\ldots\ldots\ldots\ldots \textbf{(1)}$$

In Section II of the present paper a method is described for finding the form of $f(I,E)$ as a function of I. This was carried out for three values of E (6, 12, and 18 MeV.), and for two particular cylindrical sodium iodide crystals, one $2\cdot5$ cm. long by $2\cdot5$ cm. diameter, and the other $7\cdot5$ cm. long by 4 cm. diameter, referred to as the 1 in. crystal and the 3 in. crystal respectively.

The function $g(p,aI)$, which includes the effect of statistical variation in the number of electrons ejected from the photosurface by equal flashes of light and in the multiplication of the tube, was assumed, for any value of aI, to be a Poisson distribution function of p with mean value proportional to aI. Section III

describes a measurement of the ratio of the standard deviation to the mean of the distribution as a function of the mean value.

In Section IV experimental pulse-height distributions are compared with the function $h(p,E)$ derived from (1).

II. CALCULATION OF $f(I,E)$

The function $f(I,E)$ could in principle be calculated analytically from the theory of γ-ray absorption and of energy loss of electrons. However, the multiplicity of the processes involved renders this approach impracticable. Instead, it was decided to use the " Monte Carlo " technique, which Wilson (1952) has applied to the similar problem of shower production in lead.

The principle of the method consists of following through a large number of individual events, the course of each being determined by a statistical process according to predetermined probabilities. The crystal is divided into equal intervals of length, and an incoming γ-quantum is followed through each of these sections in turn. Sets of curves are drawn, representing the probabilities of the processes of γ-ray absorption, and also of bremsstrahlung from electrons, in each interval. These are shown in Figures 1 (a)-(d), the intervals being 0·5 cm. in sodium iodide. For a particular value of energy, the difference in ordinate between adjacent curves represents the probability of the particular process occurring per interval and giving products of the stated energy. For this purpose the energy is here " quantized " into increments of 0·5 MeV. The total ordinate of the figure represents unit probability per interval.

An incident γ-quantum of a particular energy E is followed using Figure 1 (a), important areas of which are shown magnified in Figures 1 (b) and 1 (c). Its energy fixes a value of the abscissa. A four-figure number is taken from a published list of random numbers (Tippett 1927 ; Kendall and Smith 1939). If the ordinate corresponding to that number falls within a set of curves, then the appropriate absorption process is said to have taken place in the first interval, giving products whose energies can be read from the curves. On the pair production curves, the numbers give the energy E_n of the negative electron, so that the positron has energy

$$E_p = E - E_n - 1 \text{ MeV.}$$

The Compton-scattering curves are marked with the energy of the electron E_n, so that the scattered photon has energy

$$E' = E - E_n.$$

If the ordinate does not fall within a set of curves, then the quantum is said to have passed through the interval without absorption, and a further random number is taken to decide its fate in the next interval. The quantum is followed through in this manner until it either is absorbed or escapes from the crystal.

The positive and negative secondary electrons are treated using Figure 1 (d), where the curves represent differential radiation probability per 0·5 cm. in sodium iodide. An electron is followed through successive intervals, in each of which its energy is reduced by ionization, and further energy may be radiated as bremsstrahlung, until either its energy is reduced to zero or it escapes from the crystal.

The bremsstrahlung photons are then followed using Figure 1 (a), as also are positron annihilation quanta, until the shower is complete. The total energy I lost by electrons in ionization is then noted. This process is repeated for a large number of incident quanta of the same energy E, and the resultant histogram of the number of events giving ionization energy I against I is then an estimate of the form of f(I,E) for the particular value of E in that crystal.
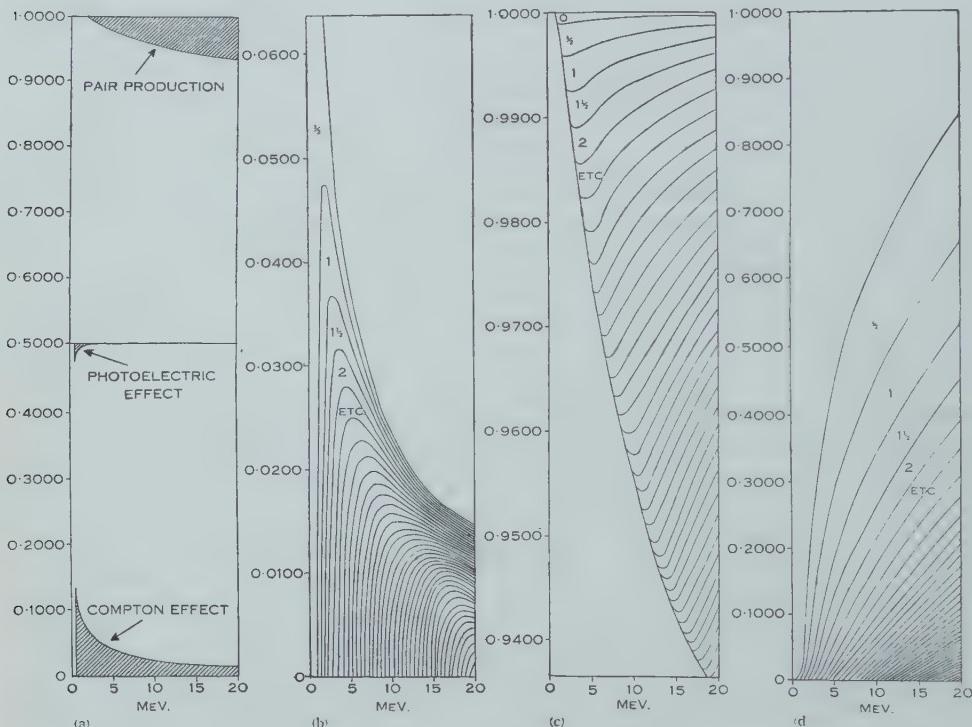
Fig. 1.—Monte Carlo probability curves.

(a) Probability curves for photon absorption.
(b) Enlargement of Compton section in (a).
(c) Enlargement of pair production section in (a).
(d) Probability curves for radiation from electrons.

Wilson carried out the random selection of ordinate by spinning a cylinder, around which the curves were wrapped. The ordinate was determined by a fixed pointer when the cylinder came to rest. Since such a cylinder would need to be spun tens of thousands of times during this investigation, a long time would be spent waiting for the cylinder to stop. Also, precautions would be necessary against bias in the cylinder. It was decided that the use of random number tables to give the ordinate directly would be more rapid and more reliable.

Figures 1 (a)-(d) were derived mainly from data given by Rossi and Greisen (1941) and Davisson and Evans (1952).

The curve for the photoelectric absorption probability in Figure 1 (a) was derived from Figures 13–17 in the paper by Davisson and Evans.

The Compton-scattering curves in Figure 1 (b) were derived from the Klein-Nishina formula (Heitler 1936). For incident photons with energy E, the differential cross section per electron for scattered quanta with energy E' is

$$\frac{d\sigma}{dE'} = -\frac{2\pi\mu g}{E^2}\left\{\frac{\mu^2}{E'^2} + \frac{1}{E'}\left(E - 2\mu - \frac{2\mu^2}{E}\right) + \frac{\mu}{E}\left(2 + \frac{\mu}{E}\right) + \frac{E'}{E}\right\}$$

over the range

$$E\left(1 + \frac{2E}{\mu}\right)^{-1} \leqslant E' \leqslant E,$$

where μ is the rest energy of an electron and g is a constant equal to $3\cdot92 \times 10^{-26}$ cm.². When this was plotted as a function of E' for various values of E, the areas in each $0\cdot5$ MeV. range gave the spacing of the curves.

The curves of Figure 24 in the paper by Davisson and Evans were replotted as a function of negative electron energy E_n, and the areas under them measured in each $0\cdot5$ MeV. interval. These areas gave the spacing of the pair production curves in Figure 1 (c).

The spacing of the curves in Figure 1 (d) was derived from Figure 7 in the paper by Rossi and Greisen. This gives the differential radiation probability for fast electrons per radiation length in lead, and, since this quantity varies little between the elements, it was taken to apply to sodium iodide directly. A radiation length in sodium iodide is almost exactly $2\cdot5$ cm. and so $0\cdot5$ cm., the interval of length used, is one-fifth of a radiation length.

In the present paper it is assumed that the γ-rays enter the crystal parallel to the axis and that all primary processes take place in the forward direction, i.e. all the products proceed forward in the same line. For pair production by 6 MeV. photons, the solid angle into which the positron and electron are emitted is of the order of $0\cdot1$ sterad. ; as the energy increases, the solid angle decreases. For the photoelectric effect, which is of importance only at energies of $0\cdot5$ and 1 MeV., the angle which the photoelectrons make with the direction of the γ-rays is of the order of $10°$. For Compton scattering, one extreme case is that which occurs when the electron receives the maximum energy possible ; it then moves off in the forward direction, while the scattered photon, whose energy is approximately $0\cdot25$ MeV., moves backward. At the other extreme, the photon moves on practically undeflected, while a low energy electron is scattered at $90°$. Thus, when products of Compton encounters move in directions far removed from that of the incident photon, their energies are small and they are soon captured. In all these cases there is little error introduced if it is assumed that all the products move off in the forward direction.

The electrons do not continue to move in a straight line but are multiply scattered until eventually their motion becomes completely random. Since it would be very difficult to treat this accurately, it was decided to adopt the approximation found satisfactory by Wilson (1951, 1952). He assumed that an electron continues to move in a straight line until its energy falls to some critical

D

energy, after which it remains at the same point in the crystal executing a
" tight random motion " until its energy is reduced to zero.   Using the data
given by Wilson, a curve was drawn giving the relation between initial energy
and critical energy for sodium iodide.

Since a bremsstrahlung photon emitted by an electron moves in nearly the
same direction as the electron, the radiation from electrons above their critical
energy travels longitudinally through the crystal, whilst that emitted by
randomly moving electrons is isotropic.   The probability that this isotropic
radiation is captured in the crystal was calculated for the case when it originates
from the centre.   The values obtained for various photon energies were assumed
to hold even when the radiation does not come from the centre of the crystal,
thereby underestimating somewhat the amount of bremsstrahlung which
escapes.   The further assumption was made, that, once the isotropic radiation
is captured, its energy is completely dissipated as ionization within the crystal,
thus further underestimating the energy which escapes.

A disadvantage of the technique just described is that any detail is lost
which is as fine as or finer than the increments into which energy is quantized.
Using increments of $0 \cdot 5$ MeV., this is particularly serious when the incident
energy is only 6 MeV., and yet it would not be practicable to carry out the whole
calculation with smaller increments.   The difficulty can be overcome by breaking
down the situation into its component processes, ascertaining in which of these
the fine detail is introduced and using smaller increments in those sections of the
calculation alone.   In the present problem, primary photoelectric effect and
Compton scattering do not introduce any fine detail, and neither does primary
pair production up to the annihilation of the positron.   It is the two $0 \cdot 5$ MeV.
annihilation quanta, either of which may or may not be absorbed, which dis-
tribute into three sharp peaks $0 \cdot 5$ MeV. apart the pulses from those primary
pairs which otherwise give all their energy to the crystal.   For this reason, it
was decided to quantize energy in $0 \cdot 1$ MeV. increments when absorption of
annihilation quanta was considered.

Therefore, in the procedure actually adopted for finding the form of $f(I,E)$,
the Monte Carlo technique was not used for the primary processes.   Instead,
the primary processes giving products in each energy range were assumed to
occur in each section of the crystal with frequencies proportional to the calculated
probabilities.   The life histories of the products for a number of examples in
each case were then followed through by the Monte Carlo method, and the
results combined appropriately.

The length of the larger crystal was divided into 20 sections, the first 10
each $0 \cdot 25$ cm., and the second 10 each $0 \cdot 5$ cm. long.   For each of the three
energies, 6, 12, and 18 MeV., 200 photons were considered, each of which produces
a pair in the first section of the crystal.   From Figure 1 (c) the number of pairs
was found in which the electron has zero energy, the number in which the
electron has $0 \cdot 5$ MeV., etc.   The development of each shower was recorded in
the manner shown by the typical diagrams in Figure 2.   Taking each pair in
turn, the motion first of the electron and then of the positron was followed
through the crystal.   From the data of Heitler (1936) and of Halpern and Hall

(1948), the ionization loss was found to be close to 2 MeV./cm. in sodium iodide and to be practically independent of electron energy above 1 MeV. Therefore 1 MeV. was subtracted from the electron energy on crossing each 0·5 cm. interval. Whether or not an electron also loses energy by radiation in an interval is determined by a random number from the tables. With the electron energy as the abscissa and the random number as the ordinate, a point is determined on the figure. If it lies above the topmost curve, no energy is lost by radiation. Otherwise, the electron loses the energy associated with the region in which the point lies, and a photon of the same energy appears in that section of the crystal.
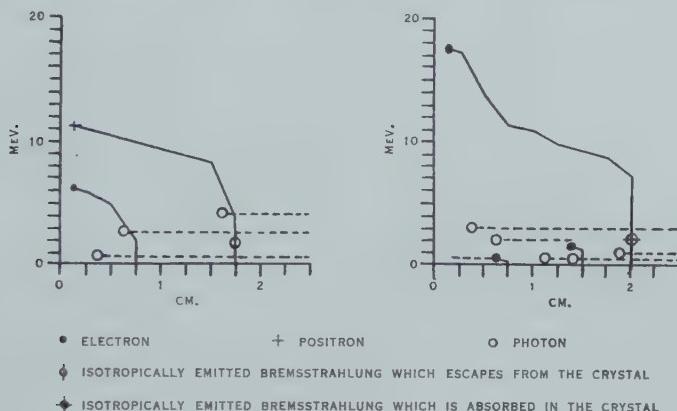
Fig. 2.—Typical 18 MeV. showers commencing in the first interval of the 1 in. crystal.

This process is continued until the energy of the electron falls below the critical energy. Then the processes occurring in each interval of length are still determined as above until the electron energy falls to zero, the only difference being that the electron moves no further forward in the crystal. The motion of the positron is treated similarly.

There exist now on the diagram two types of photons : those radiated by an electron moving in a straight line and those radiated during its random motion. The former are followed through each remaining interval in turn. In each case a new random number determines whether or not the photon is absorbed, and if so, in what way. The photon may continue unabsorbed and eventually escape from the crystal, or it may form a pair, be Compton scattered, or eject a photoelectron, in which cases all the products are followed until they are completely absorbed or escape. Those photons radiated during the random motion of the electron require only one random number to determine whether they are absorbed or not, as mentioned above. If a positron-electron pair is formed as a secondary or tertiary process, the quanta resulting from the annihilation of the positron are treated in a similar way to isotropically emitted bremsstrahlung. We now have a set of diagrams representing the consequences of a number of photons producing pairs in the first section of the crystal.

Considering just the first 10 intervals, i.e. the 1 in. crystal, it can be seen that by removing the last section and everything that occurs in it, we have the consequences of a number of photons producing pairs in the second interval of the crystal, and by removing the last two sections, of those forming pairs in the third interval, and so on. Thus, to obtain a number against energy distribution for the number of photons forming a pair anywhere within the crystal against the energy I lost by them in ionization in the crystal, we first obtain separate number-energy distributions for the energy lost by ionization for initial pair formation in each section. These relations are then combined by adding them
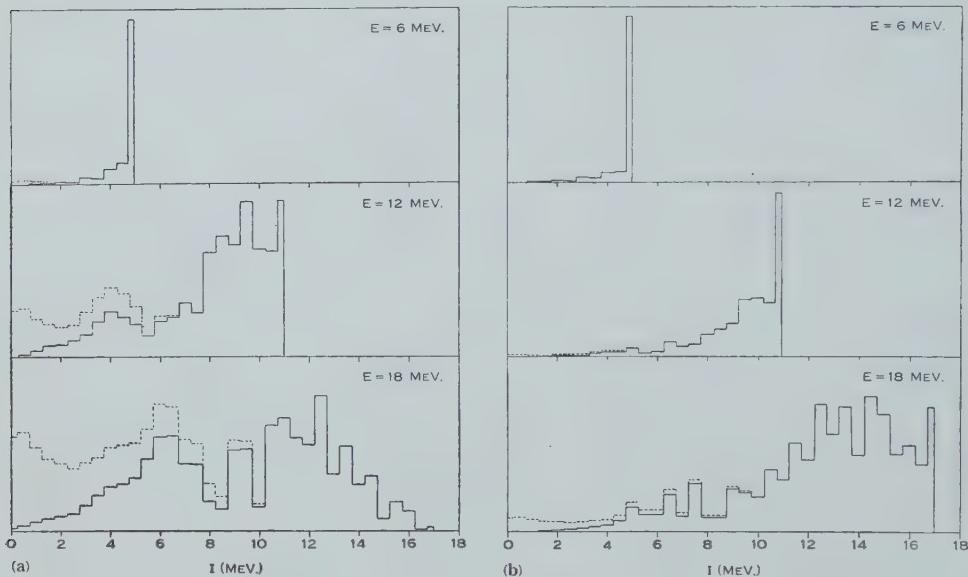
Fig. 3.—Distributions of ionization energy from primary pair production.
(a) The 1 in. crystal. (b) The 3 in. crystal.

in the ratio of the number of photons absorbed by pair production in each section, as determined by values of the absorption coefficient. Except for some modification arising from the two sizes of interval, the same treatment was applied for the 3 in. crystal. Thus we obtain the distributions of energy lost by ionization in the crystal by quanta initially absorbed by pair production, except for the annihilation radiation from the positrons which has yet to be considered. These are shown as the dashed curves in Figure 3. The full curves exclude the contribution of those showers resulting in the escape of the positron.

The probability of annihilation of a positron before coming to rest is small (Heitler 1936), and it is here assumed that the two annihilation quanta are emitted in opposite directions along a line randomly oriented in space. A numerical calculation was undertaken to determine the energy loss to the crystal of such pairs of quanta with points of origin spread uniformly through the volume of the crystal. The annihilation quanta were assumed to be of $0 \cdot 5$ MeV., and energy was considered in discrete steps of $0 \cdot 1$ MeV. The energy loss

distributions for single quanta obtained at one stage of the calculation were consistent with those found by Maeder and Wintersteiger (1952a, 1952b). The histograms obtained for the energy loss probability distributions for pairs of annihilation quanta are shown in Figure 4.

The distributions in Figure 3 were then corrected for additional absorption of annihilation energy by spreading out the full curves numerically according to Figure 4 and then replacing the area between the curves, thus completing the study of the energy loss of those incident quanta which are absorbed by the pair production process.
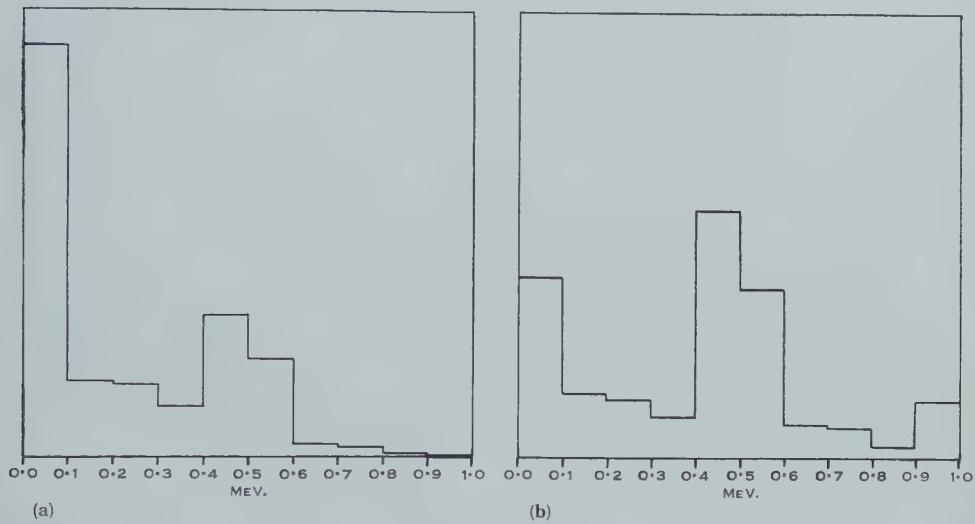
Fig. 4.—Distributions of ionization energy from annihilation quanta.
(a) The 1 in. crystal. (b) The 3 in. crystal.

Since at the energies considered primary photoelectric absorption is rare (with absorption coefficient about 1 per cent. of that of either of the other processes at 6 MeV.), it is here neglected, and it remains now only to consider those quanta initially absorbed by the Compton process.

As in the case of pair production, 200 incident quanta were considered to undergo Compton scattering, the relative numbers giving products of each energy being proportional to the calculated probabilities. The products were then followed by the Monte Carlo method until each shower was complete, in the manner previously outlined. The resultant probability distributions for ionization loss from initial Compton scattering are shown in Figure 5.

The complete distributions were then obtained by adding the pair and Compton distributions in the ratio of their relative probabilities. The final curves giving the form of $f(I,E)$ for all the cases considered are shown in Figure 6.

In some of the distributions there appears a spurious peak in the low energy region. This is a result of the particular assumption made as to multiple scattering. Consider a number of pairs with the same total energy starting from the same point in the crystal. These will all move forward approximately
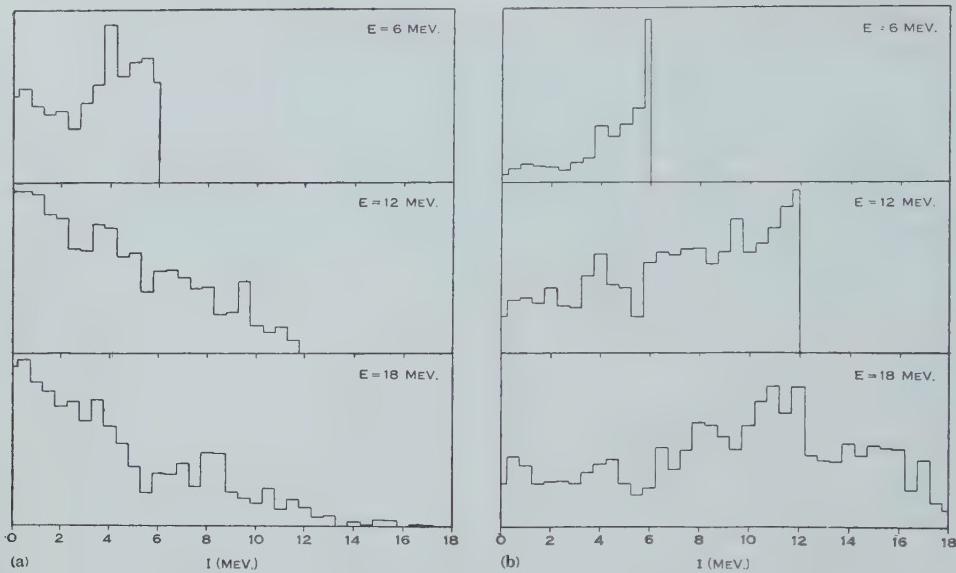
Fig. 5.—Distributions of ionization energy from primary Compton scattering.
(a) The 1 in. crystal.   (b) The 3 in. crystal.
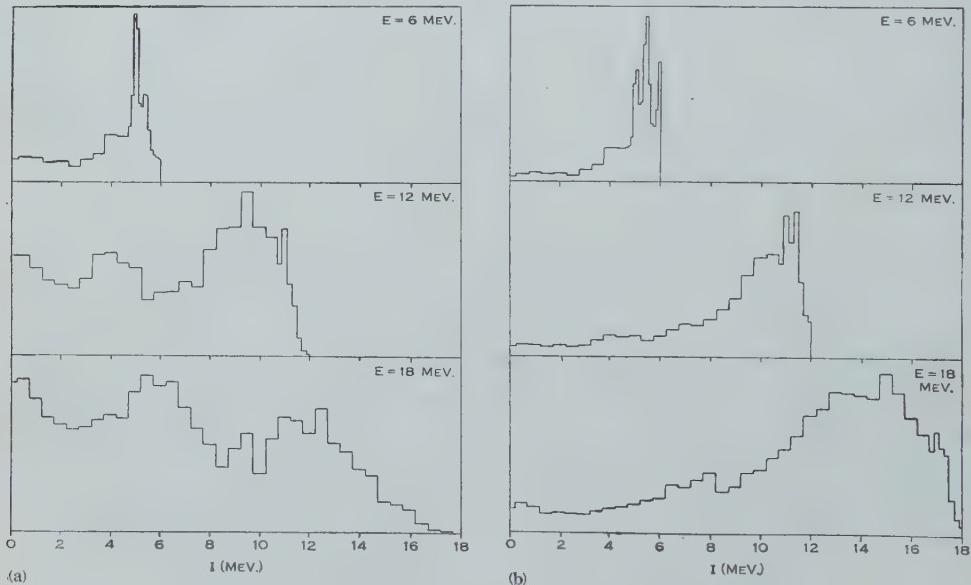
Fig. 6.—Distributions of ionization energy from all processes.   (Final forms of $f(I,E)$.)
(a) The 1 in. crystal.   (b) The 3 in. crystal.

the same distance losing a certain amount of energy by ionization, whereafter
they virtually come to rest and the remainder of their energy is dissipated in
the same part of the crystal.   Thus, if only the contribution of electrons which
finally escape from the crystal is considered, it will be seen from the above that,

because of the particular assumption made, the number as a function of energy lost will be almost constant up to a certain energy, whence it will drop quickly to zero.   Now in reality this would not be so, because the distribution from this cause begins to fall gradually from the low energies and extends further on the high energy side.   In constructing the curves in Figure 10 these spurious peaks were smoothed over in accordance with this explanation.

III. Determination of $g(p,aI)$

The EMI type 5311 photomultiplier tube used in this experiment has a photocathode sensitivity of 40 μA./lumen and a gain of 10^7 at 160 V. per stage (manufacturer's figures).   In order to measure directly the broadening of the pulse-height distribution which is introduced by statistical effects in the tube, it was necessary to expose it to flashes of light of constant integrated intensity and to measure the spread in height of the resulting output pulses.

The statistical spread is a function not only of the integrated light intensity but also of the wavelength of the light and the manner in which the intensity varies with time.   Thus, ideally, it would be necessary to measure the spread for light flashes of colour and shape identical with those from the phosphor with which the photomultiplier is to be used.   However, by inserting at the collecting electrode an integrating time constant long compared with both the test pulses and the scintillation pulses (and ensuring that the multiplier is never overloaded), the effect of differences in pulse characteristics can be minimized so that only the integrated light intensity remains important.   In the present work an integrating time constant of 4 μsec. was used, which is long compared with the light flashes from thallium-activated sodium iodide, which have a decay time of 0·25 μsec.

The method used to produce short flashes of light was to allow a beam of light reflected from a rotating mirror to pass across the photocathode of the photomultiplier.   If the mirror rotates at f rev./sec., the angular velocity of the reflected beam is $4\pi f$ rad./sec., and with the photocathode l cm. from the mirror the reflected beam passes across it at a linear speed of $4\pi fl$ cm./sec.

The experimental arrangement, shown in Figure 7, was set up in a service tunnel beneath one of the University buildings.   A 35 mm. slide projector P, with an object lens of 16 in. focal length, was used to provide the optical system, the slit being cut in a piece of fogged film.   It was found necessary to feed a direct current to the projector lamp to avoid A.C. ripple in the light intensity. The rotating mirror, RM, was mounted on a spindle which was inserted in the chuck of a high speed drill, D.   The back of the mirror, as well as the mounting, was painted black, leaving a reflecting area 2 cm. by 1·5 cm.   The reflected beam could pass through a hole in a black curtain, C, into the otherwise dark section of the tunnel, T.   At the other end of the tunnel was a fixed plane mirror, FM, to return the beam to the photomultiplier, PM, which was situated near the black curtain, so that the tunnel length was used twice.   The path length traversed by the reflected beam from the rotating mirror to the photocathode was 107 m.

The rotational speed of the mirror was found with a " Stroboflash " unit to be 11,100 r.p.m. after warming up, with short-term variations of less than 0·2 per cent., so that the linear speed of the slit image over the photocathode was 24·2 cm./μsec. Two slits were used, giving images 5 cm. and 10 cm. wide at the photocathode, which is 2·5 cm. in diameter. Thus the flashes of light gave pulses of total length 0·3 and 0·5 μsec. Both these times are short compared with the integrating time constant.

The output pulses from the photomultiplier were fed through a preamplifier with a gain of −1 to the A_1 amplifier of a Cossor model 1035 oscilloscope and thence to a single channel pulse-height analyser built to the design of Farley (1953) and to a type 1009A discriminator unit used as monitor. Pulse heights at the analyser were between 15 V. and 45 V., and the channel width was about 0·8 V. No correction was made for the broadening introduced by this channel width.
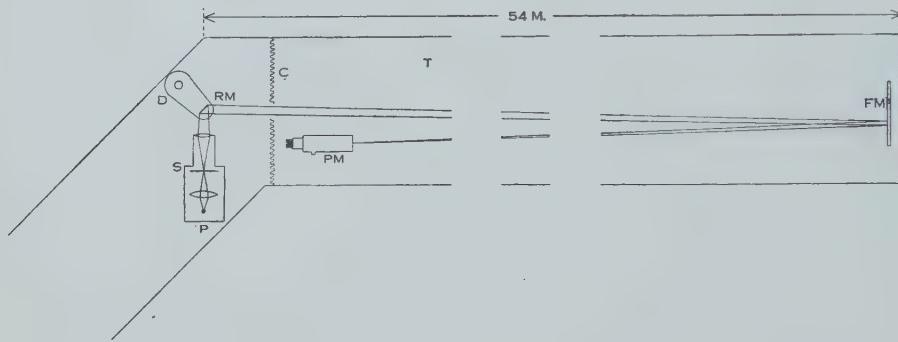
Fig. 7.—Experimental arrangement for measurement of statistical spread.

A number of runs back and forth over the pulse-height distribution was made for each of several lamp voltages with both slits, using total multiplier voltages of 930, 1040, and 1150 V. The gain of the amplifier was set in each case to give output pulse heights within the range of the analyser. The two runs shown by the points in Figure 8 illustrate the spread obtained in the two extreme cases, for the least and for the greatest integrated light intensity per flash. It can be seen that these distributions are consistent in shape with the Poisson distribution (curves), as were all others.

The mean value and standard deviation of the pulse-height distribution were calculated in each case, and the value of (standard deviation/mean) was plotted against √(multiplier gain/pulse height at collector), the result being shown in Figure 9. The multiplier gain was measured relative to that at 930 V., and the pulse height at the collector was found by dividing the mean of the pulse-height distribution in volts by the measured amplifier gain. The fact that the points are well fitted by a straight line through the origin indicates that the fractional spread of the distributions is inversely proportional to the square root of the number of electrons entering the multiplier, a characteristic feature

of the Poisson distribution, and therefore that sources of spread in pulse heights from uniform light flashes other than statistical variations are unimportant. Furthermore, there is no great dependence of the spread on multiplier voltage in the range used.
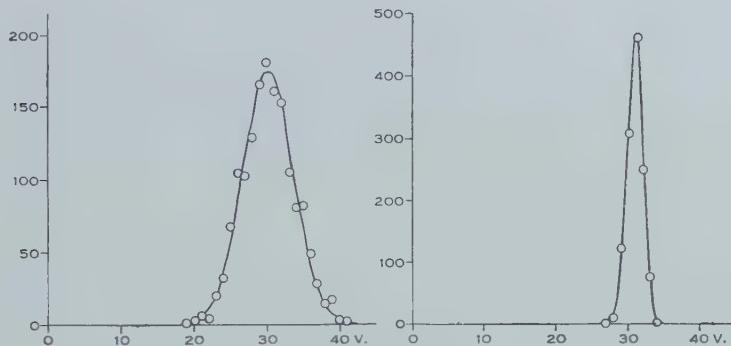
Fig. 8.—Pulse-height distributions from light flashes of constant integrated intensity.
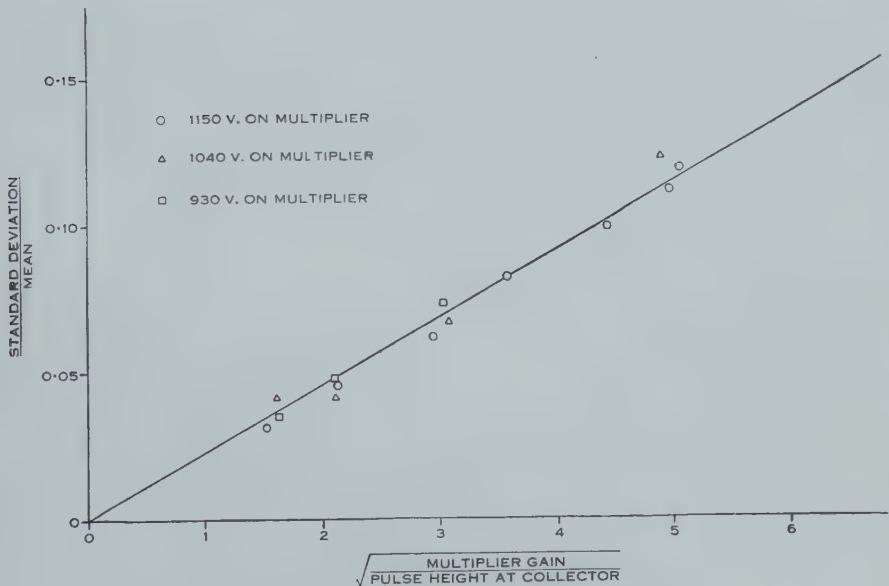
Fig. 9.—Results of determinations of statistical spread.

The slope of the line drawn through the points in Figure 9 was determined by the least mean square method. Taking the case of a 1 V. pulse at the collector (which in our apparatus had a measured capacity to earth of 20·0 μμF.) with 1150 V. on the multiplier, this value of the slope gives the ratio of the standard deviation to the mean to be 0·047, corresponding to an effective number of electrons subject to statistical variation of 453.

IV. COMPARISON WITH EXPERIMENT

The thallium-activated sodium iodide crystals used in this experiment were obtained from the Harshaw Chemical Company, Cleveland, Ohio. They were polished with successively finer grades of abrasive paper, and finally a cloth, while being kept wet with " Nujol ".

For mounting on the photomultiplier tube, the axis of which was vertical, the smaller crystal (2·5 cm. long by 2·5 cm. diameter), with aluminium foil wrapped round its curved and upper flat surfaces, was immersed in " Nujol " in a small glass beaker which was stood directly on the photocathode, with a few drops of " Nujol " added to couple optically the two glass surfaces. Thus the small crystal was mounted almost immediately on the photocathode.

The larger crystal (7·5 cm. long by 3·8 cm. diameter) was a composite one consisting of three separate crystals, each 2·5 cm. long by 3·8 cm. diameter, placed together on the one axis. As it was too large to mount directly on the photocathode, it was sealed into a " Perspex " block, the narrow space between crystal and " Perspex " being filled completely with " Nujol ". The block was tapered down to a flat circular face 1 in. in diameter, which was placed against the photocathode with a layer of glucose for optical coupling, so that the centre of the crystal lay on the axis of the photomultiplier about 4 in. from the photocathode, and the crystal axis made a right angle with the photomultiplier axis. The external surface of the " Perspex " had all sharp corners bevelled and, except for the area facing the photocathode, was covered by a layer of magnesium oxide about 1 mm. thick deposited by holding it in the smoke of burning magnesium ribbon. Mechanical support was achieved by a light frame of $\frac{1}{8}$ in. brass rod, into which the " Perspex " block was inserted before the deposition of the magnesium oxide ; the frame was then bolted to the steel plate on which the photomultiplier was mounted.

Stray light was excluded from the system by a box of thin brass sheet, which enclosed crystal, photomultiplier, high tension bleeder, and preamplifier and was screwed down to the steel plate and sealed with a rubber gasket. The vacuum tubes of the preamplifier were wrapped in black insulating tape to block light from their heaters.

The type 5311 photomultiplier, preamplifier, amplifier, and pulse-height analyser were those described in Section III. The range of the analyser was increased to 150 V. by the addition of further variable bias to the threshold amplifier, and the channel width used was 2 V.

The γ-radiation was obtained by bombarding thick fluorine, boron, and lithium targets with protons from a 700 kV. electrostatic generator. The targets were deposited on copper backings by evaporating lead fluoride, by pressing amorphous boron, and by evaporating lithium hydroxide, respectively. The bombarding energy used on all three targets was 510 keV. The γ-radiation entered the 1 in. crystal 9 in. from the target, and the 3 in. crystal 5 in. from the target.

The experimental pulse-height distributions so obtained are shown as histograms in Figure 10. Because of the greater distance of the 3 in. crystal from

the photocathode and the intervening thickness of " Perspex ", the light collection efficiency from it was much less than from the 1 in. crystal. Using the same amplifier gain throughout, the pulses from both crystals were brought within the range of the analyser by using a multiplier voltage of 766 V. with the 1 in. crystal and 1040 V. with the 3 in. one.

The distributions for the complex spectra resulting from the three nuclear reactions were obtained from Figure 6 as follows :

$F^{19}(p,\alpha\gamma)$. The predominant γ-ray energy is 6·1 MeV., with 4 per cent. of 7·1 MeV. (Hornyak *et al.* 1950). The curve for 7·1 MeV. was assumed to
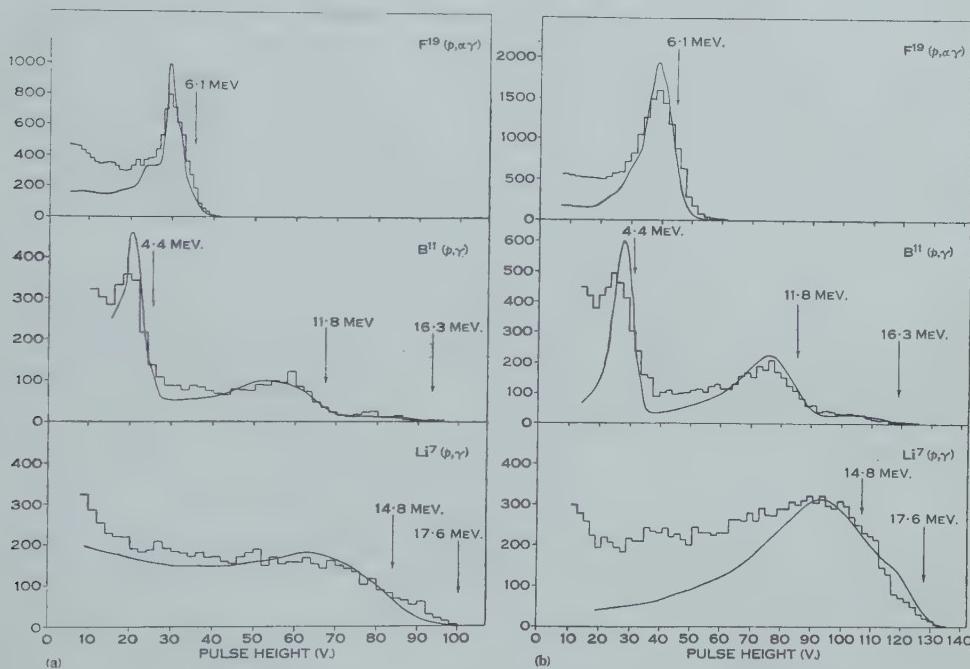
Fig. 10.—Results for nuclear spectra.
Histograms—experimental pulse-height distributions ; smooth curves—predicted forms of $h(p,E)$.
(a) The 1 in. crystal. (b) The 3 in. crystal.

have the same shape as the 6 MeV., and was drawn so that the area under it was 4 per cent. of the area under the distribution for 6·1 MeV., and the two were then added to give the distribution for this reaction.

$Li^{7}(p,\gamma)$. The γ-ray energies are 17·6 and 14·8 MeV. The distribution for 14·8 MeV. was found by interpolation between those for 12 and 18 MeV. According to Walker and McDaniel (1948) the intensities of the two components are in the ratio 2 : 1. The ratio in which the components are absorbed differs slightly from this because of the difference in absorption coefficients for the two components in sodium iodide. No allowance was made for the natural spread in energy of the 14·8 MeV. component.

$B^{11}(p,\gamma)$. The relative intensities of the three components at 4·4, 11·8, and 16·3 MeV. are 4 : 4 : 1 (Walker 1950). The 4·4 MeV. distribution was

taken to be that for 6 MeV. displaced along the energy axis, and the $16 \cdot 3$ MeV. estimated by interpolation between 12 and 18 MeV. The components were then added in the ratios given above, slightly modified to allow for the different absorption coefficients.

From the positions of the peak in the fluorine distribution it was found from the results of Section III that the effective number of electrons subject to statistical variation for each MeV. of ionization in the crystal was 335 with the 1 in. and 48 with the 3 in. crystal. Since some of the variation is introduced in the multiplier stages, these correspond to somewhat higher numbers of electrons from the photocathode entering the first stage. From these data, the integration necessary to obtain the form of $h(p,E)$ was carried out graphically.

The predicted distributions, shown as continuous curves in Figure 10, were scaled horizontally to fit the peak in the experimental fluorine distribution for each crystal. The resulting factors are $5 \cdot 70$ V./MeV. for the smaller and $7 \cdot 26$ V./MeV. for the larger crystal. The vertical scales were chosen in each case to fit the experimental histogram.

V. Discussion of Results

It can now be seen that the broadening of the peaks in pulse-height distributions for homogeneous γ-radiation above 6 MeV. is explained as the effect of bremsstrahlung escaping from the crystal and taking with it an appreciable and variable fraction of the incident photon energy.

The histograms of $f(I,E)$ in Figure 6, neglecting statistical fluctuations and the spurious peaks due to the multiple scattering assumption, correspond to the pulse-height distributions obtainable if the spread due to electron statistics in the photomultiplier were very small. These show that at 12 and 18 MeV. the increase in crystal size from 1 in. to 3 in. considerably reduces the tail at the low end but leaves the peak still quite broad.

The increase in crystal size increases the reabsorption probability not only of bremsstrahlung but also of positron annihilation radiation. At 6 MeV. this results in a change from one prominent peak (events in which neither annihilation quantum is captured) to three peaks of comparable height which would be difficult to resolve experimentally. Thus for many purposes the smaller crystal would be preferable at this energy, unless it were possible to use a crystal so much larger again that only the peak due to capture of both quanta is prominent.

Another feature seen in these histograms is that, at energies where bremsstrahlung loss is important, the peak of each histogram corresponds to some energy less than the incident γ-energy, and the nearer to it the larger is the crystal. The cut-off at the high end corresponds in each case to the incident energy.

The forms of $h(p,E)$ are compared with the experimental pulse-height distributions in Figure 10. The important features in the comparison are the abscissa values at which peaks and cut-offs occur in the $B^{11}(p,\gamma)$ and $Li^7(p,\gamma)$ spectra, the position of the $F^{19}(p,\alpha\gamma)$ peak having been used for calibration.

These are such as to confirm the results of the Monte Carlo calculation of the effect of bremsstrahlung escaping from the crystal. As regards the actual shapes of these peaks, it appears that in each case the experimental peak is spread out even more than predicted. This can be explained by the fact that, wherever approximations were necessary in the Monte Carlo calculation, they were made in such a direction as to underestimate the amount of bremsstrahlung which escapes, so that the predicted curves exhibit energy resolution better than attainable.

The width of the peak for the $F^{19}(p,\alpha\gamma)$ radiation is also greater than predicted. This could be due to the channel width of the analyser, non-uniform sensitivity over the photocathode surface, and inequality of light collection efficiency between different parts of the crystals. In the case of the 3 in. composite crystal, any variation in fluorescence efficiency between the individual crystals would add to the spread.

The lower ends of the pulse-height distributions diverge upward from the predicted curves. This can be explained as the effect of radiation scattered into the crystal from matter outside it. No collimation was used, and no particular care was taken to minimize the amount of scattering material in the neighbourhood of the target and crystal. In particular, the γ-radiation entered the 3 in. crystal through a depth of about $\frac{3}{8}$ in. of " Perspex ".

Following a suggestion of Hofstadter and McIntyre (1950), some workers have reported improved energy resolution with scintillation counters above 1 MeV. by using three crystals in triple coincidence to isolate those pulses due to events in which a positron is annihilated in the main crystal and both annihilation quanta escape from it (Johansson 1950 ; Bair and Maienschein 1951). At the energies considered in the present paper, the form of $f(I,E)$ for a detector of this type would be practically that given by the full lines in Figure 3, since the creation of a positron by other than the incident photon does not often occur. For 6 MeV. radiation, the low energy tail is small, but it increases both with increasing energy and with decreasing crystal size. Comparing Figure 3 with Figure 6, it is seen that the spreading due to bremsstrahlung is less than in the single crystal case, because the average energy of the secondary electrons is less. At 18 MeV., however, the spreading has become so great that there is little to be gained by using the triple coincidence technique and, at higher energies still, all advantage disappears because pair production becomes virtually the only primary process.

Therefore with γ-radiation above 10 MeV. it is a fundamental limitation to the energy resolution of scintillation counters that a fraction of the incident energy is lost to the crystal by bremsstrahlung in each detecting event and that this fraction differs widely between events. The only way to overcome this appears to be to increase the size of the crystal so that the escaping fraction is kept small. Thus, while at lower energies the minimum crystal size is determined by the necessity to ensure that electron ranges are small compared with crystal dimensions, above 10 MeV. the crystal dimensions would need to be large compared with the mean free path of the bremsstrahlung photons. In

sodium iodide, this is about 4 cm. for 1 MeV. photons, increasing to 8 cm. for 5 MeV. and then decreasing very slowly at higher energies.

Nevertheless, sodium iodide crystals of the sizes we have used are still useful up to 18 MeV., having the properties that

   (i) the efficiency of detection, as shown in Table 1 for photons entering parallel to the crystal axis, is good ;

  (ii) the energy of a component of the radiation can be estimated by the position not of the peak but of the end-point of the distribution ;

 (iii) pulses from the highest energy component can be isolated by a simple discriminator set above the end-point of the distribution due to the component next below.

TABLE 1

PERCENTAGE DETECTION EFFICIENCY OF SODIUM IODIDE CRYSTALS

Photon Energy (MeV.)	1 in. Crystal (%)	2 in. Crystal (%)
6	27	62
12	30	66
18	33	70

VI. Acknowledgments

The authors express their appreciation to Professor L. H. Martin for his encouraging interest. Thanks are due to the Warden of the Melbourne University Union for use of the tunnel, and to Dr. F. J. M. Farley for details of his pulse analyser circuit. Finance for equipment was provided by C.S.I.R.O.

VII. References

Bair, J. K., and Maienschein, F. C. (1951).—*Rev. Sci. Instrum.* **22** : 343.

Davisson, C. M., and Evans, R. D. (1952).—*Rev. Mod. Phys.* **24** : 79.

Farley, F. J. M. (1953).—*J. Sci. Instrum.* (in press).

Garlick, G. F. J., and Wright, G. T. (1952).—*Proc. Phys. Soc. Lond.* B **65** : 415.

Halpern, O., and Hall, H. (1948).—*Phys. Rev.* **73** : 477.

Heitler, W. (1936).—" The Quantum Theory of Radiation." (Oxford Univ. Press.)

Hofstadter, R., and McIntyre, J. A. (1950a).—*Phys. Rev.* **79** : 389.

Hofstadter, R., and McIntyre, J. A. (1950b).—*Phys. Rev.* **80** : 631.

Hornyak, W. F., Lauritsen, T., Morrison, P., and Fowler, W. A. (1950).—*Rev. Mod. Phys.* **22** : 291.

Johansson, S. A. E. (1950).—*Nature* **166** : 794.

Kendall, M. G., and Smith, B. B. (1939).—" Tables of Random Sampling Numbers." Tracts for Computers, No. 24. (Cambridge Univ. Press.)

Maeder, D., and Wintersteiger, V. (1952a).—*Phys. Rev.* **87** : 537.

Maeder, D., and Wintersteiger, V. (1952b).—*Helv. Phys. Acta* **25** : 465.

Rossi, B., and Greisen, K. (1941).—*Rev. Mod. Phys.* **13** : 240.

Seitz, F., and Mueller, D. W. (1950).—*Phys. Rev.* **78** : 605.

TAYLOR, C. J., JENTSCHKE, W. K., REMLEY, M. E., EBY, F. S., and KRUGER, P. G. (1951).-
    Phys. Rev. **84** : 1034.
TIPPETT, L. H. C. (1927).—" Random Sampling Numbers." Tracts for Computers, No. 15
    (Cambridge Univ. Press.)
WALKER, R. L. (1950).—Phys. Rev. **79** : 172.
WALKER, R. L., and McDANIEL, B. D. (1948).—Phys. Rev. **74** : 315.
WILSON, R. R. (1951).—Phys. Rev. **84** : 100.
WILSON, R. R. (1952).—Phys. Rev. **86** : 261.

THE DISTRIBUTION OF RADIO BRIGHTNESS OVER THE SOLAR DISK AT A WAVELENGTH OF 21 CENTIMETRES

I. A NEW HIGHLY DIRECTIONAL AERIAL SYSTEM

By W. N. Christiansen* and J. A. Warburton*

Summary

A new aerial system of very high resolving power has been designed for use in determining the distribution of radio brightness across the solar disk at a wavelength of 21 cm. Thirty-two aerials with paraboloidal reflectors are evenly spaced in an east-west direction over a distance of about 700 ft., and are connected by a branching system of balanced open-wire transmission lines to a receiver. The aerial system produces multiple beams each 3′ of arc wide and spaced 1·7° apart. The rotation of the Earth causes one after another of the aerial beams to scan the disk of the Sun. The record obtained from the radio receiver gives a one-dimensional brightness distribution over the solar disk.

I. INTRODUCTION

The poor resolving power of radio aerial systems used in radio astronomy is in marked contrast with the high resolution obtainable with optical telescopes, and is the main limiting factor in astronomical studies by means of radio waves. Until recently even a relatively large astronomical object like the Sun could not be resolved by the aerial systems in use.

The radio-frequency emission from the Sun has been studied intensively in recent years, and several different components of the radiation have been discovered. One of these is constant over months or years, and has been identified with thermal radiation from the undisturbed or " quiet " Sun. A determination of the brightness distribution of this component would be valuable in providing information about physical conditions in the solar atmosphere.

Another component, prominent at decimetre wavelengths, has been called the " slowly varying " component of the solar radiation. Its variations show a good correlation with visible sunspot area.

Radio observations during solar eclipses, commencing with those of Covington (1947), have shown that small areas of enhanced radio brightness often occur in the vicinity of sunspots, and there is little doubt that these areas are the source of the slowly varying component of the solar radiation. Eclipse observations are rare, however, and do not reveal the changes with time of these regions of enhanced brightness. Insufficient information is available,

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

therefore, to test theories regarding the nature of these regions. Their presence at decimetre wavelengths produces difficulties in interpreting eclipse records, and has prevented a clear determination of the brightness distribution of the quiet-Sun component over the solar disk. The most useful eclipse observations have been made at the highest frequencies where the variable component of the solar radiation is small.

At decimetre wavelengths frequent observations are necessary to separate the quiet-Sun and slowly varying components, hence an aerial of very high resolving power is required.

II. Aerial Systems of High Resolving Power

For a detailed study of the solar disk, the aerial system should have a beam width much less than the angular diameter of the Sun, say 1/10 of this or about 3′ of arc. An aerial system 1000 wavelengths wide is required to achieve this resolution. At a wavelength of 2 dm., this is about 700 ft. It would be impracticable to construct a steerable aerial system of such a size and of conventional design.

High resolution in one angular coordinate requires the aerial to be large in only one direction and is therefore very much easier to achieve than a pencil beam. The aerial beam is then fan-shaped and in radio astronomical work this has the great advantage that the aerial need not be steered, but instead the Earth's rotation may be used to sweep the aerial beam across the object to be studied. If the aerial system has the form of a uniform strip, it produces a single main beam. The Sun would pass through this beam only once each day. If the system has the form of a non-uniform strip, then more than one main beam may be produced. An extreme form is the two-aerial interferometer which has a large number of beams of equal amplitude. The latter arrangement, however, is not useful for finding the brightness distribution across an extended object like the solar disk because the spacing between adjacent beams is only double the width of each beam.

A modified form of the two-aerial interferometer was used by Stanier (1950), who varied the spacing between the two aerials and from the different interference patterns derived a brightness distribution for the quiet Sun at a wavelength of 60 cm. Undetected areas of enhanced brightness on the disk may produce serious errors in the results obtained by this method.

III. Principle of a New Aerial System

For an aerial system in which there is a number of elements arranged at uniform intervals along a straight line, the polar diagram takes the form of multiple beams having a separation from each other which is inversely proportional to the spacing between adjacent elements. Such an arrangement is suitable for studying the Sun, as beams of the required narrowness can be produced and sufficient elements used so that the Sun will be in no more than one aerial beam at any time. Such a system is analogous to a diffraction grating or, less directly, to a multiple-beam optical interferometer. A maximum response is obtained when the optical path to adjacent aerials is the same or

E

differs by an integral number of wavelengths.   The directivity of systems such as
this (in terms of received power) is given by

$$\Phi(\theta) = \frac{\sin^2 Np}{N \sin^2 p}, \quad \dots\dots\dots\dots\dots\dots\dots \quad (1)$$
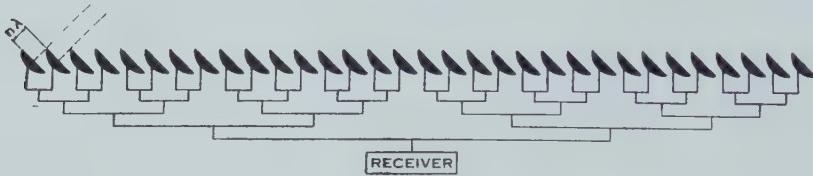
Fig. 1.—General arrangement of the aerial system and transmission lines.   A
maximum is obtained when n is integral.
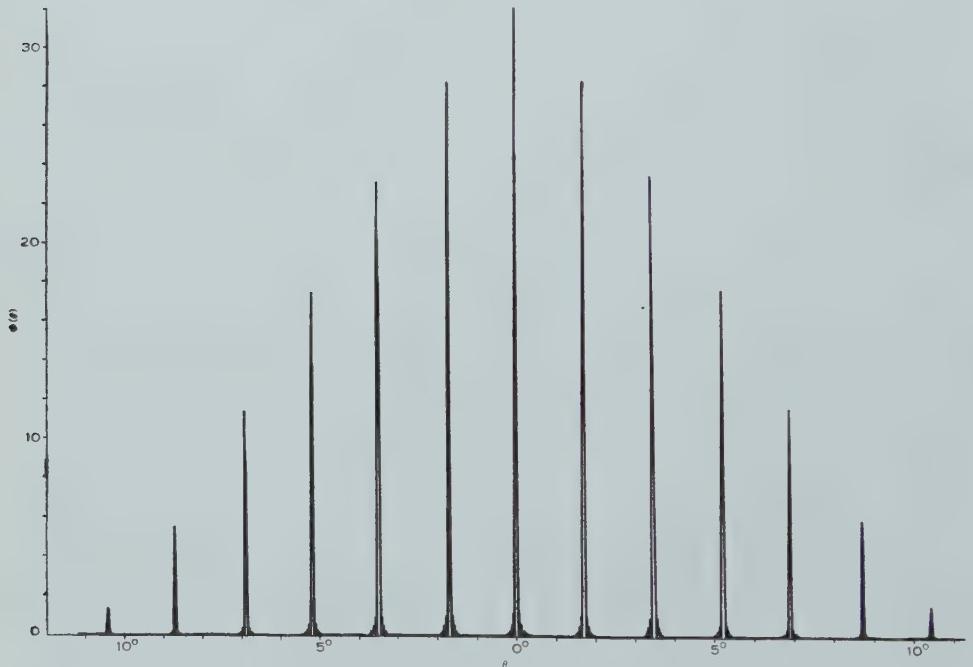
Fig. 2.—Directional diagram of the 32-aerial interferometer.

where $p = \pi d \sin \theta / \lambda,$

$\Phi(\theta) =$ the power received by the system from a point source, relative
   to the power received by one of its elements,

$N =$ number of elements in the system,

$d =$ spacing between adjacent elements,

$\lambda =$ wavelength,

$\theta =$ angle between the normal to the system and the direction of the
   ray.

A system of this type operating at a wavelength of 21 cm. has been built in Sydney. It consists of 32 elements arranged over a distance of about 700 ft. in a straight line, roughly in an east-west direction. A schematic diagram of the system is shown in Figure 1.

The fan-shaped beams have a calculated beam width of 2·9′ of arc between half-power points and there is a spacing of 1·7° between adjacent beams. A representation of the beam system is shown in Figure 2. The envelope to the family of beams corresponds to the directivity pattern of the individual aerials of the system. Since these individual aerials can be steered, the envelope is movable and can follow the Sun as it passes through the large number of aerial beams during the course of a day. (Both the beam width and the angle between adjacent lobes increases, however, as the angle made with the normal to the line of the aerial system increases.)
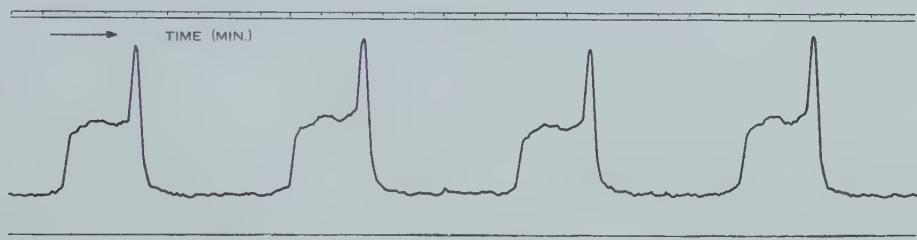
Fig. 3.—A typical record of the passage of the Sun through several aerial beams.

The radio-frequency energy from the aerials is amplified in a radio receiver and the output fed to a recording milliammeter. As the Sun passes through a beam of the aerial, the deflection of the meter at any instant is proportional to the energy being emitted from the narrow strip of the solar disk in the aerial beam at that instant. A series of records which gives the one-dimensional brightness distribution across the solar disk is obtained. A typical record is shown in Figure 3.

IV. A Modified Form of the Aerial for Circular Polarization

In the arrangement described above, the polarization of all the aerials is horizontal, and that of the complete system is the same. A modified form of the system has also been used. Adjacent aerials are polarized in mutually perpendicular planes and the complete system resolves the incident radiation into circularly polarized, right- and left-hand, components. The aerials are connected in the way shown in Figure 4, so that there is a 90° change in polarization between one aerial and the next, the change being progressive from one end of the system to the other. For directions in which there is a difference of $\frac{1}{4}$ wavelength in the optical path to adjacent aerials, circularly polarized radiation falling on the aerials will induce currents in the aerials which will be either additive or will have a zero resultant, depending on the sense of rotation of the waves. If the optical paths to adjacent aerials differ by $\frac{3}{4}$ wavelength, the effect is reversed ; the sense of polarization which produced additive currents at a

path difference of $\frac{1}{4}$ wavelength now produces zero resultant, while radiation with the opposite sense produces additive currents. The system will now be examined in more detail.

Consider a wave from a distant source, incident on the aerial system, with components of magnitude ε_x parallel to aerials 1, 3, 5, etc., and ε_y parallel to 2, 4, 6, etc. The response v_x of the aerial system to ε_x is given by

$$v_x = \alpha \varepsilon_x \sin(\omega t - 2\pi l_1/\lambda + \delta_1).f(\theta). \quad \dots \dots \dots \dots \quad (2)$$

Similarly, the response v_y to ε_y is

$$v_y = \alpha \varepsilon_y \sin(\omega t - 2\pi l_2/\lambda + \delta_2).f(\theta), \quad \dots \dots \dots \dots \quad (3)$$
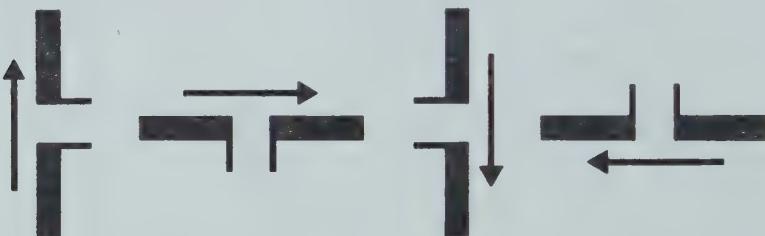
Fig. 4.—Arrangement of four aerials for reception of circularly polarized waves. Eight of these sets make up the complete system.

where α is a constant of proportionality,

$\omega = 2\pi$ times the frequency of the wave,

$\lambda =$ wavelength,

l_1, l_2 are the optical path lengths from the source to the centres of the two sets of aerials,

δ_1, δ_2 are the initial phases at the source,

$f(\theta)$ represents the directional response of either the x-plane or y-plane system of aerials and is given by

$$f(\theta) = \frac{\sin Np}{\cos 2p}, \quad \dots \dots \dots \dots \dots \dots \quad (4)$$

where N and p have the same meaning as in equation (1).

$f(\theta)$ has maxima when

$$2p = (2n + 1)\pi/2, \quad (n = 0, 1, 2, \dots)$$

i.e. twice as frequently as the system represented in equation (1).

In order to find the relative phase of v_x and v_y at these values of p, we note that

$$l_1 = l_2 + d \sin \theta,$$

since there is a separation d between the centres of the vertical and horizontal groups of aerials. Hence, at the maxima,

$$2\pi l_1/\lambda = 2\pi l_2/\lambda + 2p = 2\pi l_2/\lambda + (2n + 1)\pi/2,$$

therefore

$$v_x = \alpha N/2.\varepsilon_x \sin \{\omega t - 2\pi l_1/\lambda + \delta_1\},$$
$$v_y = \alpha N/2.\varepsilon_y \sin \{\omega t - 2\pi l_1/\lambda + (2n + 1)\pi/2 + \delta_2\}.$$

If ε_x and ε_y are equal and the phase difference $(\delta_2 - \delta_1)$ between the two components at any point along the path is equal to $\pi/2$, i.e. the wave is right-hand circularly polarized, then the sum

$$v_x + v_y = 2v_x, \text{ if } n \text{ is odd,}$$
$$= 0, \text{ if } n \text{ is even.}$$

If $\delta_1 - \delta_2 = \pi/2$, i.e. the wave is left-hand circularly polarized, then

$$v_x + v_y = 2v_x, \text{ if } n \text{ is even,}$$
$$= 0, \text{ if } n \text{ is odd.}$$

Hence, for any direction of maximum response for the aerial, one circularly polarized component of the radiation is received, while that of opposite sense is rejected. At adjacent maxima the effect is reversed.

The above treatment applies strictly only at the maxima, but the rejection of one circularly polarized component is practically complete over the whole of any beam, because the relative phase of ε_x and ε_y changes slowly with θ. Even at the edge of the beam, i.e. near the first minimum, the phase has been changed only by about $10°$, so that the sensitivity of the system to one component is 100 times greater than to the other.

As the Sun passes through one aerial beam after another, the receiver responds to right-hand and left-hand circularly polarized components of the radiation alternately. If the radiation over the whole Sun is linearly or randomly polarized, then successive records are equal in amplitude and similar in shape. If the radiation from a particular part of the Sun is circularly polarized, the response to this appears only on alternate records. With elliptically polarized radiation, or a mixture of circularly and randomly polarized radiation, the amplitude of the response would be different on successive records.

This modified form of the equipment has been used principally in investigating the radiation which originates near sunspots.

V. Design

(a) Aerial System

The 32 aerials (shown in Plate 1 (a)) which form the system are arranged in a straight line at intervals of 23 ft. In setting up the system a maximum error of 1 in., or $\frac{1}{8}$ wavelength in the placement of aerials was allowed. A later survey showed that this had not been exceeded.

Individual aerials, which have a beam width of about $10°$, consist of a dipole radiator and reflector placed at the focus of a paraboloidal reflector, 66 in. in diameter. Each is mounted on a polar axis and capable of rapid position change in hour angle. A simple system of pegs and holes in a circular plate allows accurate positioning of each aerial. When in use the positions of the aerials are changed by hand approximately each $\frac{1}{4}$ hr.

Each dipole aerial and reflector is supported on a pair of parallel flat strips of metal attached to the paraboloidal disk as shown in Plate 1 (b). These strips serve, in addition, as a balanced open-wire transmission line. At the lower end, $\frac{1}{4}$ wavelength from a shorting strip, a short length of flexible polythene-insulated balanced transmission line connects the aerial to the main transmission line.

(b) Transmission Lines

The aerials are connected to the receiver through a branching system of transmission lines in the way indicated in Figure 1, so that there are equal lengths of transmission line between each aerial and the receiver. With such an arrangement the system is insensitive to small changes in frequency or to changes in the length of the feeder lines caused by temperature variations, because each aerial is affected equally. This is most important when the length of transmission line is greater than 500 wavelengths.

The transmission lines are of the balanced open-wire type. The reasons for the choice of this type of line, unconventional at such a high frequency, are that it is easily accessible for connections, matching, and testing and can be designed to have low attenuation at small expense.

The pairs of conductors consist of hard-drawn copper wire, 0·16 in. in diameter, with a spacing between centres of 0·37 in. The characteristic impedance of the line is approximately 180 ohms. Each pair is under a tension of about 500 lb. wt. and is supported every 23 ft. on polystyrene insulators. To preserve the correct spacing between the conductors, small polystyrene spacers are placed between the lines at intervals of about 6 ft. Both the supporting insulators and the spacers are arranged in pairs separated by $\frac{1}{4}$ wavelength in order to reduce reflections in the line.

The transmission-line attenuation resulting from resistive losses in the conductors was calculated to be approximately 1·1 db./100 ft. of line. Calculation indicated that the dielectric and radiation losses would be much less than this. Measurements over straight sections of the transmission line showed that the attenuation was approximately 1 db./100 ft. The overall loss in the line was found to be 6 db. The additional loss apparently takes place in the various junctions, matching sections, and short-circuited sections along the line.

Reflections in the line are small. It was found that a standing-wave ratio of 1·1 could be preserved. Small discontinuities in the line have been compensated, usually by a suitably placed single polystyrene insulator.

The branching structure of the transmission line, shown in Figure 1, has been effected by running five pairs of lines one above the other. At a junction a short pair of vertical conductors connects adjacent lines. The unwanted part of the transmission line is effectively disconnected by shorted quarter-wave sections, and matching is effected by use of a quarter-wave matching section formed from two plates of metal. A photograph of such a section is shown in Plate 1 (c). A very simple way of adjusting this was adopted. The correct position for the matching section was found by use of a standing-wave indicator, the two metal plates were hung on the line and their spacing varied until a correct match was indicated, and, finally, the plates were soldered in position.

In the setting up of the system, power was fed into the transmission line at the signal frequency, and matching was effected by means of a standing-wave indicator. This consisted of a pair of pick-up probes made from brass rod embedded in a block of polystyrene. The latter was shaped to fit onto the two-wire line, and had a handle by which it could be slid along the line. The

probes were connected to a shielded pair of coaxial lines which led, through a balance-to-unbalance transformer, to a single coaxial line terminating at a radio receiver. The D.C. output from the detector of this receiver was taken back through long meter leads to a milliammeter close to the place of measurement.

The aerials were matched individually to the transmission lines. They were then short-circuited and the electrical centre of the transmission line joining each pair of aerials was determined from the positions of voltage minima on the line. This position was checked by length measurements. (In all cases there was negligible difference between the positions of the electrical and physical centres.)

At this central point, a short vertical section of line was used to connect the first transmission line to a second line situated vertically below, as shown in Plate 1 (c). The two-to-one mismatch at the junction was corrected by means of a quarter-wave matching section. The same procedure was adopted at each branch.

(c) The Receiver

A radio receiver is connected to the centre of the transmission-line system through a balance-to-unbalance transformer. The receiver is of conventional superheterodyne design.

At the input is a radio-frequency switch consisting of a rotating condenser and transmission-line sections, which at a frequency of 25 c/s. connects the receiver alternately to the aerial system and to an effectively infinite length of coaxial transmission line. Following the switch is a coaxial cavity tuned to the signal frequency and leading to a crystal detector. A line-tuned heterodyne oscillator is also coupled to the detector, which in turn is connected to a 30 Mc/s. amplifier of 4 Mc/s. bandwidth. The 30 Mc/s. amplifier is followed by a detector, a 25 c/s. amplifier, and a phase-sensitive detector. The output of the latter is connected to a recording milliammeter.

The receiver is calibrated each day by means of a thermal noise generator, which has an output corresponding to a temperature approximately 100 °C. above ambient. This temperature difference produces a change in the receiver output which is approximately equal to that produced by the passage of the Sun through the aerial beam. The noise fluctuations on the receiver record have an r.m.s. value corresponding to a temperature of about 1 °C.

VI. Observational Technique
(a) Daily Recording

The aerial system has been in operation since February 1952. For most of the time it has been used in its original form, i.e. with all dipole aerials in the same plane of polarization. Records, such as those shown in Figure 3, are obtained each day over a period of about 2 hr. around midday, when the resolving power of the aerial is greatest. Outside this time, two unfavourable effects become significant. The first is the increase in beam width which occurs as the projected length of the system diminishes in a direction normal to the Sun's direction. The second is a distortion which results from the increasing number

of wavelengths difference in the optical paths to different unit aerials.   This means that frequencies in different parts of the receiver pass-band have their maxima displaced from each other.   The second effect may be reduced by decreasing the receiver bandwidth.   A bandwidth of 0·5 Mc/s. can be used if records are required outside the normal hours, but this has the adverse effect of causing an increase in the noise fluctuations on the records.
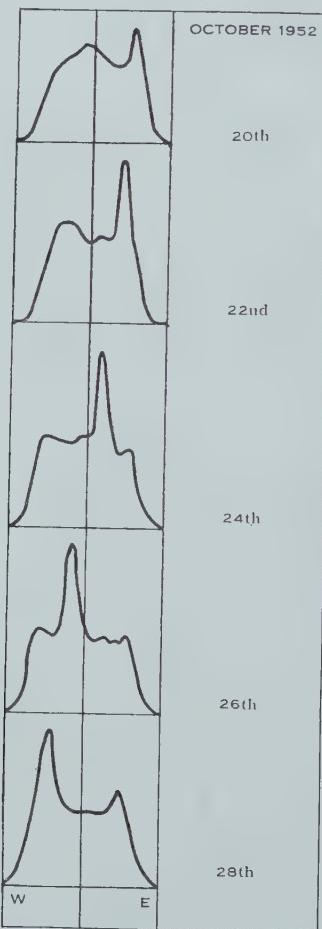
Fig. 5.—A succession of daily records showing the one-dimensional brightness distribution over the solar disk.

The effect of noise fluctuations in the records is diminished by averaging a number of records for each day.   A series of " averaged " daily records is shown in Figure 5.   No convenient way of measuring the beam width of the aerial system has yet been found.   However, on several occasions when there have been small areas of great radio brightness on the disk, an upper limit to the beam width could be determined.   This is approximately 3·4′ of arc, which is a little greater than the calculated value for a point source.

(b) *Intensity Calibration*

The calibration of the receiver in terms of the effective temperature at the input, in the way described earlier, is not completely adequate to detect changes in gain of the system, because it does not indicate any changes that may occur in the aerial system. A comparison is made each day, therefore, between the total amount of energy received from the Sun as indicated, on the one hand, by the integral of the recorded curve of solar disk brightness obtained with this equipment, and on the other hand, that indicated by different equipment which measures daily the energy received from the whole solar disk. (The comparison equipment operated at 25 cm. instead of 21 cm., but results indicate that no significant error is introduced by this.) Occasional changes in aerial gain have been found, these usually being associated with faults at junctions of the transmission line. When such changes have occurred, a check is usually possible through a comparison of the heights of certain parts of the records obtained on adjacent days.

(c) *Determination of Position*

In order to make a comparison between optical and radio features of the solar disk it is necessary to know the position of a particular aerial beam with respect to the solar disk at any time. The direction of the line along which the aerial system lies is known, hence the directions in space of the various aerial beams may be calculated and compared with the known direction of the Sun at any particular time. Time marks are placed on the daily radio records, and thus the position of the Sun with respect to an aerial beam can be calculated for any part of the radio record.

Analysis over a period of many weeks showed that bright areas near the limb do not markedly increase the apparent diameter of the radio Sun. It follows that the radio Sun and optical Sun are centred on the same line in space. This fact was used to eliminate much tedious calculation, because the centre of the radio record could be taken to correspond to the centre of the solar disk, and the position of any features of the radio record determined with respect to this line.

VII. Analysis of Observations

It is not intended to analyse results in this paper, but merely to show how the observations can be used to elucidate certain solar phenomena. The daily records show characteristic peaks which change in position from day to day, moving at approximately the same rate as sunspots, i.e. they indicate local areas of enhanced brightness which rotate with the Sun. In addition, the records appear to show a lower level below which the brightness does not fall, and this presumably represents the radiation from the quiet Sun. In order to study the two components it is necessary to separate them. The first step is to deduce the contribution of the quiet Sun, i.e. to determine the one-dimensional brightness distribution for the thermal component. When this is done, the contributions of the areas of enhanced brightness are represented by projections above the quiet-Sun level.
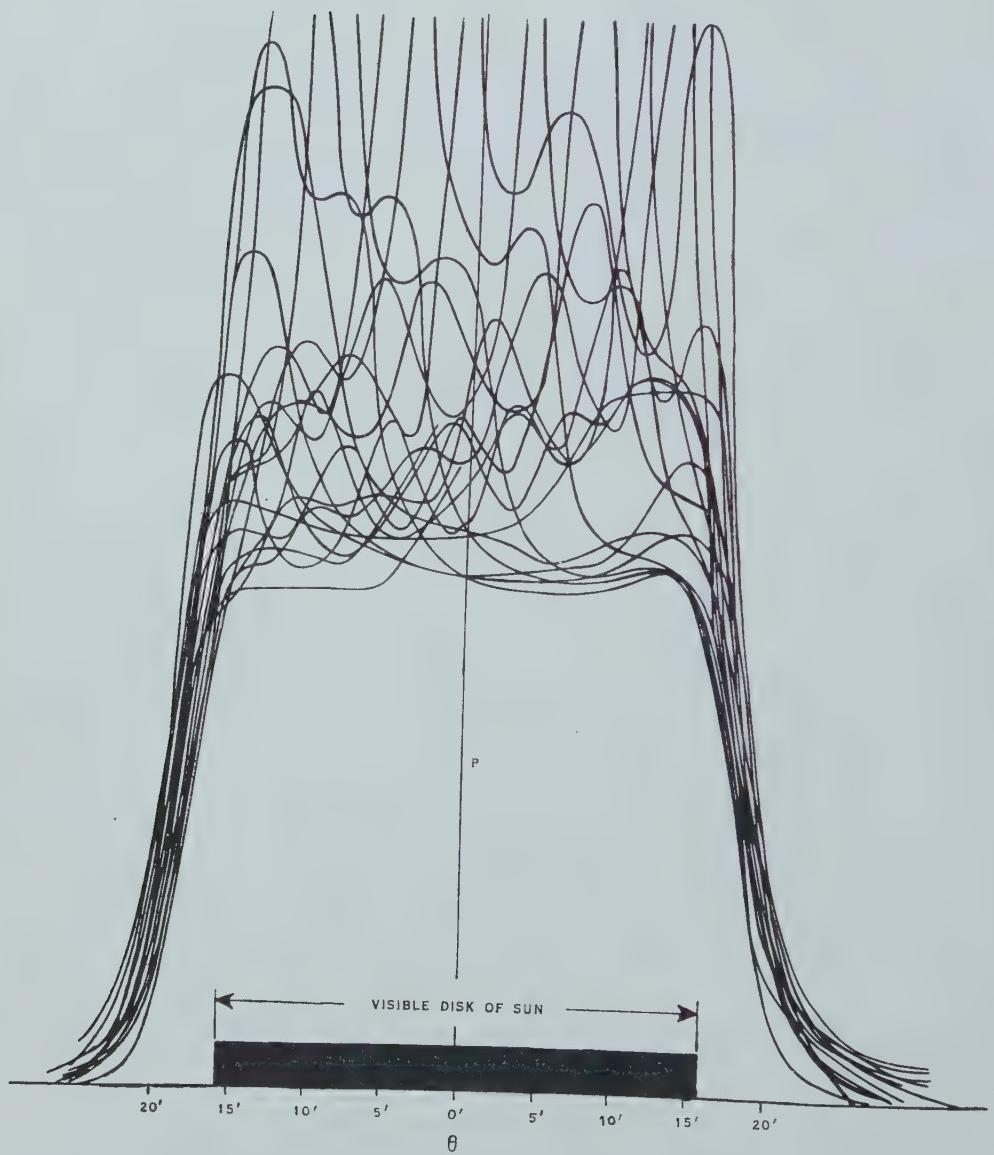
Fig. 6.—A succession of daily records superimposed to show the contribution of the quiet Sun. P is the power received, in arbitrary units. θ is the angular separation between the centre of the aerial beam and the centre of the optical disk of the Sun.

(a) Determination of the One-Dimensional Brightness Distribution of the Quiet Sun

One way in which this might be done would be to compare records in which no peaks are apparent. If these have the same shape and amplitude, one would be justified in saying that the curves represent the quiet-Sun brightness distribu-

tion.   Unfortunately cases of complete absence of areas of enhanced brightness were very rare during 1952.   (The absence of sunspots is not sufficient.)

An alternative way to determine the distribution is to plot a series of daily records on one diagram and to investigate the existence and shape of a lower envelope to the curves.   In Figure 6 a typical set of 20 superimposed records is shown.   The existence of a lower envelope to the curves is clearly seen.   The shape of this envelope will be discussed in a following paper.

(b) Study of Areas of Enhanced Radio Brightness

The positions on the disk of these areas indicated by peaks on the records change with the solar rotation.   This can be demonstrated by drawing on a Mercator's projection of the Sun for each day an arc which represents the position of an aerial beam with respect to the heliographic coordinates of the Sun when a particular peak appears on the record.   The approximate position of the area of enhanced brightness on the Sun can be found from the place of intersection of the arcs.   If this is done with different assumed values for the effective solar radius, i.e. the distance of the radio source from the centre of the Sun, then the best intersection represents the probable position in the solar atmosphere in height as well as latitude and longitude.   A comparison may then be made with optical features of the solar disk.

The energy being received from each area of enhanced radio brightness may be deduced once the quiet-Sun level has been found.   The changes in received energy from such a source depend both on real changes in total emission from the source, and on apparent changes of emission related to the position of the source in the solar atmosphere.   The possibility of separating these two effects will be discussed in a later paper.

VIII. Conclusions

The multiple-aerial interferometer principle provides a very economical means for a detailed study of the radio-emitting surface of the Sun.   The system described in this paper has been found simple to adjust and will remain in adjustment over periods of months.   The main factor in producing these desirable characteristics is undoubtedly the branching system of transmission lines which is non-critical to changes of frequency or atmospheric conditions. The construction of such a system would have been difficult if open-wire transmission lines had not been used.

The records obtained with this equipment allow the quiet-Sun and slowly varying components of the solar radiation to be separated.   In later papers the brightness distribution over the solar disk of the first of these components and the characteristics of the second component will be discussed.

IX. Acknowledgments

aerial system ; Mr. D. E. Yabsley, who provided useful criticism of this paper ; and Dr. J. L. Pawsey, whose comments and suggestions throughout the course of the work were always helpful.

X. References

COVINGTON, A. E. (1947).—*Nature* **159** : 405.

STANIER, H. M. (1950).—*Nature* **165** : 354.

RADIO BRIGHTNESS OVER THE SOLAR DISK. I
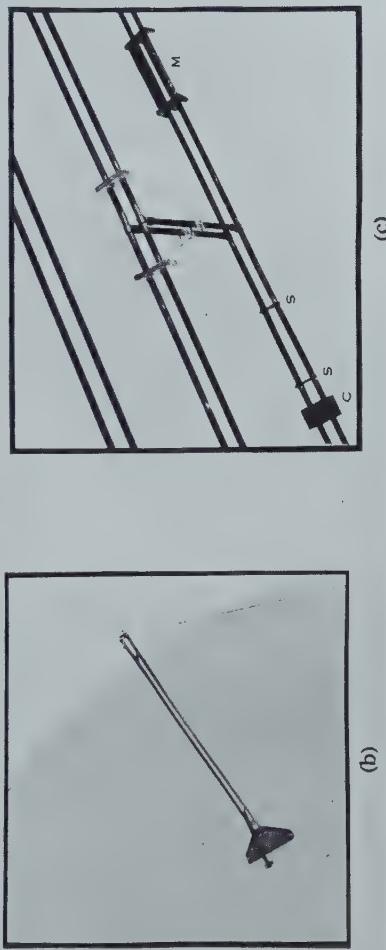
(a) The 32-aerial interferometer.

(b) Dipole-aerial and reflector.

(c) A typical junction of two transmission lines. C, a clamp; S, short circuits; M, a quarter-wave matching section consisting of two flat plates, each bent to fit on to wire conductor.

FURTHER STUDIES OF ELECTRODE PHENOMENA IN TRANSIENT ARCS

By W. R. Blevin*

Summary

High current transient arcs in air at atmospheric pressure display radial contraction of the column near the anode. An examination of melting in the anode allows calculation of the anode spot temperature.

I. Introduction

It has been known for some time that the arc discharge column contracts radially near the *cathode* to give rise to very high current densities at that electrode (Cobine and Gallagher 1948 ; Froome 1948, 1949, 1950 ; Somerville and Blevin 1949). The phenomenon has usually been studied by photographing the luminous discharge and by examination of signs of melting and tarnishing left on the electrode by the discharge.

Both of these techniques have been employed by the author to study the behaviour of arcs at the *anode*. The discharges examined had durations varying from 1 μsec. to 1 msec., and were struck between metallic electrodes in air at atmospheric pressure by moving the electrodes together till sparking occurred. High constant current pulses were obtained by discharging artificial lines. The instantaneous potential across the gap was of the order of 30 V., showing that the discharge was an arc.

II. Contraction of Arcs at the Anode

Plate 1, Figure 1, is a photograph of a 200 μsec., 50 A. arc between a copper wire cathode (top) and a polished plane aluminium anode. The photograph was taken obliquely with the shutter held open throughout the duration of the arc. Considerable radial contraction of the discharge is seen to have occurred near the anode, and a highly luminous and circular " anode spot " is apparent. A Kerr cell shutter has been used to show that such a spot is established within the first few microseconds of the life of the arc, and does not alter significantly with time (Somerville, Blevin, and Fletcher 1952). Photographs similar to Plate 1, Figure 1, have been obtained for the low melting point metals tin and lead.

* Department of Physics, New England University College, Armidale, N.S.W.

Plate 1, Figure 2, is a photograph of a similar arc struck on a silver anode, and the absence of any anode spot is typical of the higher melting point metals. Contraction of the discharge at the anode is still apparent, however. The lower half of the photograph shows the reflection of the discharge in the anode. Measurement of the cross section of the discharge at the anode for different metals shows that the current density there may be as great as 50,000 A./cm.^{2}.

III. Temperature of the Anode Spot

The temperature of the anode spot is assumed to be constant with time. A similar assumption has been made for the cathode by previous authors (Holm 1947 ; Somerville, Blevin, and Fletcher, in press) chiefly because the problem of melting is more easily solved. Now, the arc characteristics show that the total rate of supply of energy to the arc is approximately constant, while the above assumption implies that the rate of supply of heat to the anode is inversely proportional to the square root of the time. These two situations are not incompatible, because the transient arcs examined have been characterized by an expanding cathode spot, which presumably uses energy at an increasing rate, and a comparatively static anode spot, which is consistent with a state of temperature equilibrium.

It has been possible to calculate the temperature of the anode surface in the central region of the active area by observation of the extent of melting of the electrode by each transient arc. Usually the depth of melting is small compared with the anode thickness and the theory of melting in a semi-infinite solid can be applied to the physical problem with reasonable accuracy. A technique for measuring depth of melting into thick anodes has been described previously (Somerville, Blevin, and Fletcher 1952). Unfortunately this method does not give sufficiently accurate measurements, and it has been found more satisfactory experimentally to study melting by arcs struck on metallic foils, and to develop a theory of melting for that case.

Consider the case when an arc of duration T has as anode the plane surface of a thin slab of metal of thickness α. If the diameter of the discharge greatly exceeds α, there is little radial conduction of heat in the metal from the centre of the active area, and heat entering the anode in that region is conducted linearly through the slab. Assume that for the duration of the arc the active anode surface is held at a constant temperature Φ (the initial temperature of the anode being zero), and that no heat is lost from the back of the anode slab. Let V be the melting point of the anode metal.

An analogous problem has been solved by Lightfoot (1930) to give the rate of solidification of a slab of molten steel. Small changes in Lightfoot's solution allow the rate of melting of the anode in the present problem to be determined. Although the solution is not exact, it is an extremely good approximation for the region used in the present case.

The position of the level at which the temperature is V, at time t, is given by

$$x = 2k\sqrt{\varkappa t}, \quad \dots\dots\dots\dots\dots\dots\dots\dots (1)$$

where $x=$ the coordinate of depth into the anode,

$\qquad \varkappa =$ the diffusivity $=K/c\rho$,

$\qquad K=$ the thermal conductivity,

$\qquad c=$ the specific heat,

$\qquad \rho =$ the density,

and k is given by

$$\frac{\Phi}{V}=\frac{1}{1-\mathrm{erf}\ k}+\frac{\lambda}{cV}\sqrt{\pi}ke^{k^2}\ \mathrm{erf}\ k, \quad \ldots\ldots\ldots\ldots (2)$$

where $\lambda =$ the latent heat of fusion.

The temperature in the molten region $0<x<2k\sqrt{\varkappa t}$ is

$$v_1=\Phi-\frac{\Phi-V}{\mathrm{erf}\ k}\ \mathrm{erf}\left(\frac{x}{2\sqrt{\varkappa t}}\right)-\frac{V}{1-\mathrm{erf}\ k}\sum_{m-1}^{\infty}(-1)^m\left\{\mathrm{erf}\left(\frac{2m\alpha+x}{2\sqrt{\varkappa t}}\right)-\mathrm{erf}\left(\frac{2m\alpha-x}{2\sqrt{\varkappa t}}\right)\right\},$$

$$\ldots\ldots\ldots\ldots\ldots (3)$$

and beyond this region, for $2k\sqrt{\varkappa t}<x<\alpha$, the temperature is

$$v_2=\frac{V}{1-\mathrm{erf}\ k}\left[1-\mathrm{erf}\left(\frac{x}{2\sqrt{\varkappa t}}\right)-\sum_{m=1}^{\infty}(-1)^m\left\{\mathrm{erf}\left(\frac{2m\alpha+x}{2\sqrt{\varkappa t}}\right)-\mathrm{erf}\left(\frac{2m\alpha-x}{2\sqrt{\varkappa t}}\right)\right\}\right].$$

$$\ldots\ldots\ldots\ldots\ldots (4)$$

The quantity of heat entering the anode per unit area is

$$Q=\int_0^T K\left(\frac{\partial v_1}{\partial x}\right)_{x=0}\ dt. \quad \ldots\ldots\ldots\ldots (5)$$

Consider a slab of critical thickness α_0, such that melting just extends to its back surface. Then

$$Q=\alpha_0\rho(cV+\lambda). \quad \ldots\ldots\ldots\ldots (6)$$

If we now eliminate Q from (5) and (6), and substitute for v_1 from (3), we obtain

$$\frac{\lambda}{cV}=\frac{2}{\sqrt{\pi}}\frac{\sqrt{\varkappa T}}{\alpha_0}\left[\frac{\Phi/V-1}{\mathrm{erf}\ k}\right.$$

$$\left.+\frac{2}{1-\mathrm{erf}\ k}\sum_{m=1}^{\infty}(-1)^m\left\{e^{-m^2\alpha_0^2/\varkappa T}-\frac{\sqrt{\pi m\alpha_0}}{\sqrt{\varkappa T}}\left(1-\mathrm{erf}\left(\frac{m\alpha_0}{\sqrt{\varkappa T}}\right)\right)\right\}\right]-1.^*$$

$$\ldots\ldots\ldots\ldots\ldots (7)$$

Using (2) to eliminate k, we can obtain λ/cV in terms of $\sqrt{\varkappa T}/\alpha_0$ for given values of Φ/V. The solution is shown in Figure 1 for values of Φ/V in the range $1\cdot2$-10.

If the assumptions are approximately correct, Figure 1 enables the central anode spot temperature Φ to be determined. The abscissa λ/cV is a constant for a given metal and the ordinate $\sqrt{\varkappa T}/\alpha_0$ is determined by experiment. For some metals accurate measurement of α_0 is difficult, but in the cases mentioned

---

* A more elementary treatment (Somerville, Blevin, and Fletcher 1952) gives a similar result, but with the infinite series omitted. Errors due to this omission are not serious when $\sqrt{\varkappa T}/\alpha_0$ is small.

below the error is less than 5 per cent. (It is found that α_0 is independent of the arc current for values greater than 50 A., presumably because the spot diameter is then large enough for radial conduction of heat from the centre to be negligible.) Then Φ/V is read from the graph, and so Φ is known.
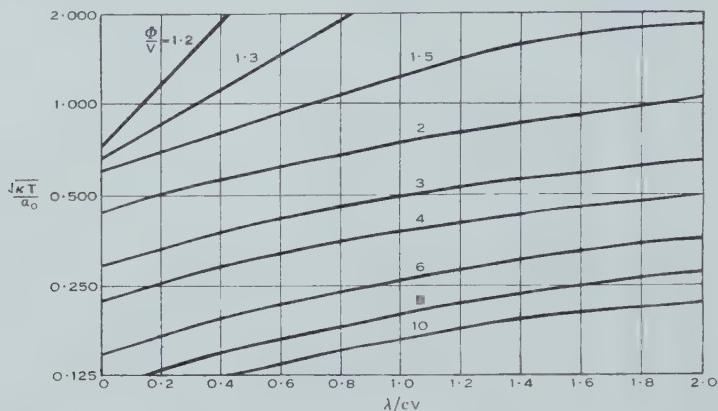
Fig. 1.—$\sqrt{\kappa T}/\alpha_0$ is shown as a function of λ/cV for values of the parameter Φ/V in the range 1·2–10.

Results for anodes of nickel, aluminium, and tin are given in Table 1. If the anode spot temperature is not constant as assumed, Φ represents a mean value. However, the results for tin suggest that the assumption of constant temperature is approximately true. For each metal Φ is well below the boiling point.

TABLE 1

VALUES OF CENTRAL SPOT TEMPERATURE

Anode Metal	Arc Duration (μ sec.)	$\dfrac{\lambda}{cV}$	$\dfrac{\sqrt{\kappa T}}{\alpha_0}$	$\dfrac{\Phi}{V}$	Spot Temperature (°C.)
Tin　..　　..　　..	27	1·06	0·22	7·5	1600
	50	1·06	0·24	6·8	1460
	250	1·06	0·25	6·5	1400
Aluminium　..　　..	250	0·45	0·63	1·9	1250
Nickel　..　　..　　..	250	0·43	0·83	1·50	2300

The calculated average surface flux of heat is consistent with the value derived from the anode drop. Our solution gives values from 2 to 9 V. for the anode fall with the above metals (assuming that all the energy of the anode fall enters the anode), and accepted values for this quantity are of the same order (von Engel and Steenbeck 1932).

It is necessary to consider whether the values of Φ in Table 1 also hold for normal electrodes of much greater thickness than α_0, for it is possible that use of a thin slab as anode may have so limited the amount of heat conducted away that the surface temperature is considerably increased. The temperature at the back of the slab at time $t=T$ gives an indication of the extent to which the use of the slab has influenced the spot temperature. With $t=T$, and $x=\alpha_0$, (4) becomes

$$\bar{v}_2 = \frac{2V}{1-\text{erf }k}\left[1-\text{erf }\left(\frac{\alpha_0}{2\sqrt{\chi T}}\right) + \sum_{m=1}^{\infty}(-1)^m\left\{1-\text{erf }\left(\frac{(2m+1)\alpha_0}{2\sqrt{\chi T}}\right)\right\}\right].$$

For tin this temperature is less than 10 °C. and the boundary $x=\alpha_0$ therefore has negligible influence on Φ. However, for aluminium and nickel the values of $\bar{v}_2$ are not much less than V, and the values of Φ for these metals are somewhat greater than the spot temperatures on thicker anodes.

Indeed, experimental examination shows that the anode surface conditions alter when very thin metal foils are used for the electrode. The area of the molten spot increases when α falls below some critical value α_1. This is partly due to increased radial conduction of heat, but photographs of the discharge show that its cross section at the anode also increases, causing a decrease in the radial contraction. A tentative explanation is that high temperatures built up in very thin anodes cause appreciable thermionic emission of electrons, which

TABLE 2

COMPARISON OF THICKNESSES α_0 AND α_1

Anode Metal	α_0 (cm.)	α_1 (cm.)
Tin	0·030	0·007
Aluminium	0·018	0·016
Nickel	0·0071	0·016

decreases the total current density. Table 2 compares values of α_0 and α_1 for 250 μsec, 75 A. arcs. For tin $\alpha_0 \gg \alpha_1$, in agreement with the above conclusion that the boundary $x=\alpha_0$ does not significantly influence the state of the surface $x=0$. However, α_0 and α_1 are of the same order of magnitude for each of the other metals, and this again suggests that the thinness of the foil in these cases is affecting the anode spot temperature. For aluminium and nickel the spot temperature on normal anodes must therefore lie between their melting points (660 and 1455 C., respectively) and the values given in Table 1.

IV. ACKNOWLEDGMENT

F

V. REFERENCES

COBINE, J. D., and GALLAGHER, C. J. (1948).—Current density of the arc cathode spot. *Phys. Rev.* **74** : 1524-30.

VON ENGEL, A., and STEENBECK, M. (1932).—" Elektrische Gasentladungen." (J. Springer : Berlin.)

FROOME, K. D. (1948).—The rate of growth of current and the behaviour of the cathode spot in transient arc discharges. *Proc. Phys. Soc. Lond.* **60** : 424-35.

FROOME, K. D. (1949).—The behaviour of the cathode spot on an undisturbed mercury surface. *Proc. Phys. Soc. Lond.* B **62** : 805-12.

FROOME, K. D. (1950).—The behaviour of the cathode spot on an undisturbed liquid surface of low work function. *Proc. Phys. Soc. Lond.* B **63** : 377-85.

HOLM, R. (1947).—Uber die Stromdichte in den ersten Momenten eines gezogenen Bogens. *Ark. Mat. Astr. Fys.* **34B** (1), Paper 8.

LIGHTFOOT, N. M. H. (1930).—The solidification of molten steel. *Proc. Lond. Math. Soc.* (2) **31** : 97-116.

SOMERVILLE, J. M., and BLEVIN, W. R. (1949).—Current densities in the cathode spots of transient arcs. *Phys. Rev.* **76** : 982.

SOMERVILLE, J. M., BLEVIN, W. R., and FLETCHER, N. H. (1952).—Electrode phenomena in transient arcs. *Proc. Phys. Soc. Lond.* B **65** : 963-70.

ELECTRODE PHENOMENA IN TRANSIENT ARCS

Fig. 1.—Radial contraction of an arc discharge at a low melting point anode. ×15.

Fig. 2.—Radial contraction of an arc discharge at a high melting point anode. ×15.

CLIMATIC CHANGE IN AUSTRALIA SINCE 1880

By E. L. Deacon*

Summary

Australian climatic data show that, for the period 1911–1950, the summer rainfall over much of the southern part of the continent was considerably greater than in the previous 30 years and, for the same season, mean daily maximum temperatures in the interior were appreciably lower. A difference in character of the annual variation of atmospheric pressure between these periods also suggests a shift of mean position of the subtropical high pressure belt.

It is tentatively concluded that, contemporaneous with the increased meridional interchange which has taken place over large parts of the northern hemisphere, a similar increase has occurred in the Australian region.

I. Introduction

The climate of large regions of the Earth has been subject to wide variation over long periods of time as has been amply shown by many lines of geological and palaeobotanical evidence. From these studies it seems that fluctuations of greatly varying intensity and time scale have occurred, ranging probably all the way from the ice age glaciations and interglacial epochs at one extreme to the year to year variations at the other. Although many theories have been advanced to explain these phenomena (see Brooks (1949) for a discussion and bibliography), none has yet achieved general acceptance and consequently much attention is being given to the lengthening series of instrumental observations in the hope that detailed study of such changes as have occurred in this period may eventually lead to an understanding of the causation of climatic variation. As yet the main emphasis is on establishing the extent of climatic change in the instrumental period and the relation of changes in any one region to those elsewhere.

Convincing evidence of an appreciable climatic trend in recent decades over much of the northern hemisphere has been put forward, and Ahlmann (1948) has reviewed much of this material together with the results of glaciological studies which have demonstrated a notable retreat and thinning of glaciers in many areas, trends which have in most cases accelerated since about the beginning of the century. The climatological evidence points to an increasing transport of heat into high latitudes by the general circulation of the atmosphere during this period, with an appreciable increase particularly in the mean winter temperatures over large areas mainly in high latitudes. A possible reversal of this trend since about 1940 is not yet fully established.

* Section of Meteorological Physics, C.S.I.R.O., Melbourne.

A similar study for the southern hemisphere is handicapped by lack of data for high latitudes and by the shorter period of instrumental observations, but, in view of the small amount so far published for this region, it was considered to be useful to make the study of Australian data presented in this paper.

The search for climatic trends in Australia, which might have resulted from a change in pattern or intensity of the general circulation of the globe, was started with the guiding idea that such a change would probably most strongly affect the summer climate, as there is at this season a very strong temperature contrast between central Australia and the Southern Ocean. Most attention has therefore been given to the summer statistics and for this purpose the months December, January, and February comprise the summer. As will be explained later, it was found convenient in much of the work to examine the differences between 30-year mean values of various climatic elements and for brevity two 30-year periods are referred to as follows :

Period 1	1881–1910,
Period 2	1911–1940.

II. Climatic Régime of Australia

Before proceeding to the changes in climatic elements since 1880 in Australia, it is necessary first to sketch very briefly the main features of the Australian climate in relation to the general circulation of the southern hemisphere, with particular attention to the more populated southern part of the continent, as the data for the north and west are too scanty for the present purpose.

The climate of Australia is largely influenced by the subtropical belt of high pressure, the axis of which moves north and south with the Sun. In winter the high pressure is sufficiently far north for the south of the continent to come intermittently under the influence of the westerlies and so have a winter maximum of rainfall. At this season the north is very dry ; Darwin, for example, with a mean annual rainfall of 60 in., receives rather less than 1 in. of this in the four months May to August.

When the high pressure belt has moved south in summer much of the north and north-east, except far inland, receives a copious rainfall under the influence of tropical low pressure areas assisted by monsoonal action. Cyclonic activity along the east coast extends the area of mainly summer rainfall southward into eastern New South Wales. The continent south of the tropic then receives little rainfall except for the region embracing the highlands of south-east Australia ; low pressure troughs between the migratory anticyclones of the high pressure belt cause this region to have a summer rainfall of similar magnitude to that in winter.

III. Air Temperature Changes

As already mentioned, a change in advective influences due to changing general atmospheric circulation would be expected most clearly to influence Australian inland summer temperatures. Climatic comparisons are commonly made in terms of some broad average such as the mean monthly or mean annual temperature. However, factors such as wind strength, cloud amount, and

wetness of the soil affect minimum temperatures inversely to the maxima, so that changes in these elements may be taking place which are not reflected to a marked degree in the *mean* temperature. For the present purpose, where evidence for change in the circulation pattern is being sought, either the mean daily maximum or the mean daily minimum temperature would provide a more sensitive and revealing index. The choice between the two is to some extent arbitrary, but the former is on the whole preferred since the high rate of turbulent mixing of the lower atmosphere in the day-time should make the maxima more representative.
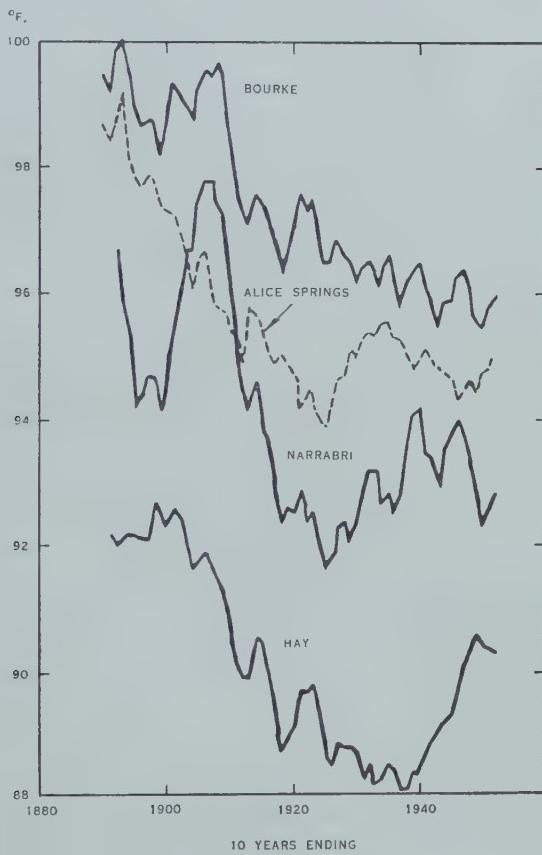
Fig. 1.—Ten-year running averages of mean summer maximum temperature.

Moving 10-year averages of mean daily maximum temperature for summer are shown in Figure 1 for some typical Australian inland stations. There are some differences in character between the graphs but all show a marked falling trend over much of the period with a levelling off or reversal in the last 10 or 15 years. This suggested studying the differences of the mean values for the two 30-year periods 1881–1910 (Period 1) and 1911–1940 (Period 2), a choice which also seemed appropriate in the light of the northern hemisphere trends. Table 1 gives these differences for the inland localities in central south and south-east
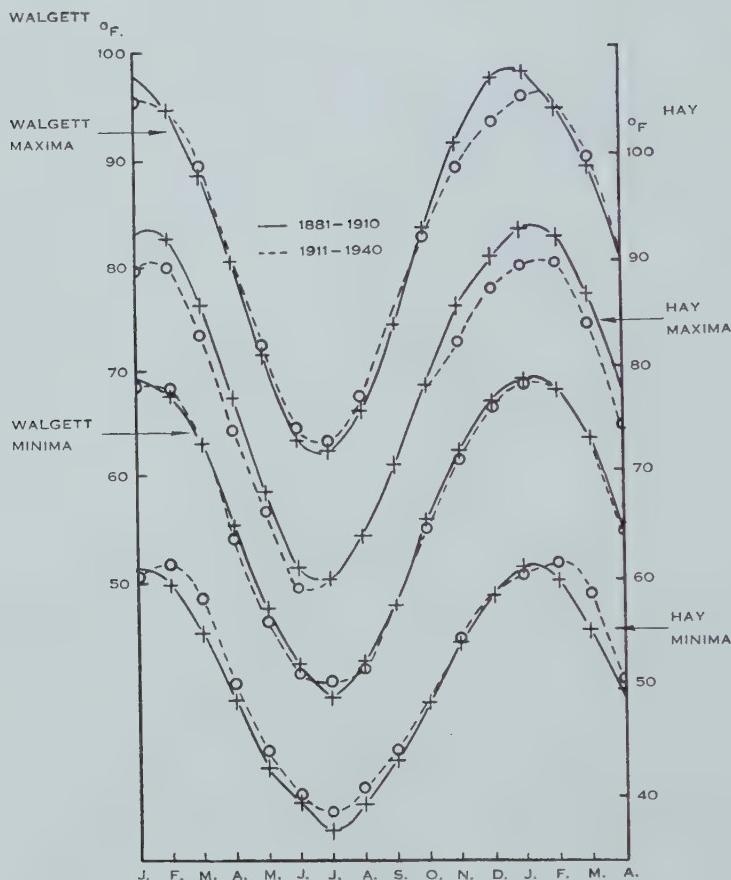
Fig. 2.—Thirty-year means of mean monthly maximum and minimum
temperatures at Hay and Walgett.

TABLE 1

DIFFERENCES BETWEEN MEAN DAILY MAXIMUM TEMPERATURES FOR THE SUMMER SEASONS OF
1881–1910 (T_1) AND 1911–1940 (T_2) FOR INLAND AUSTRALIAN LOCALITIES

Place	Lat. (°S.)	Long. (°E.)	T_1–T_2 (°F.)	Place	Lat. (°S.)	Long. (°E.)	T_1–T_2 (°F.)
Alice Springs ..	23·6	133·6	2·3	Hay	34·5	144·9	3·1
Walgett ..	30·0	148·2	2·4	Goulburn ..	34·7	149·7	0·5
Bourke ..	30·1	146·0	2·4	Albury ..	36·1	146·9	0·8
Narrabri ..	30·3	149·8	3·0	Echuca ..	36·1	144·8	1·6
Coonabarabran	31·3	149·3	2·9	Cooma ..	36·2	149·1	4·7
Dubbo	32·3	148·6	2·3	Bendigo ..	36·8	144·3	3·6
Bathurst ..	33·4	149·6	2·1	Omeo.. ..	37·1	147·6	3·0

Australia which have sufficiently complete temperature records extending back to 1881.

All these localities show lower mean summer maximum temperatures in Period 2 and the average fall of $2 \cdot 5$ °F. is comparable in magnitude to the simultaneous changes in winter temperature in north-west Europe. The good consistency of the changes suggests the cause to be mainly climatic rather than changing observational technique or exposure.

Changes in the mean maximum temperature at other times of year and in minimum temperatures at all seasons are smaller,* as may be seen from Figure 2 in which the 30-year averages of monthly mean values are shown for Hay and Walgett. As a result the annual mean temperature at these places has not changed very appreciably, the decrease from the first to the second period being only $0 \cdot 7$ °F. at Hay, $0 \cdot 4$ °F. at Walgett, and $0 \cdot 6$ °F. at Alice Springs. This result shows how little informative annual mean temperatures may be. Hanzlik (1931) has remarked on the importance of studying circulation patterns by seasonal rather than annual means. The present results suggest that, in studying temperature, it may often be necessary to go a stage further and take the maxima and minima separately, as combining the two in mean temperatures may obscure significant features, particularly in climates where clear skies predominate.

IV. Precipitation Changes

Rainfall data for south and south-east Australia were examined for evidence consistent with the trend in summer maximum temperatures. Too few records back to 1880 are available for the rest of the continent. Some rather remarkable increases in summer rainfall in the second 30-year period over that in the first period are displayed in Figure 3. It is also notable that the figures for the different localities, with but few exceptions, vary systematically over the region so that rough isopleths can be drawn with some confidence. The zero isopleth is almost coincident with the dividing line between two different rainfall régimes. The region to the north-east receives much of its rain under the influence of east coast cyclones but the summer rainfall in the southern area is largely associated with troughs of low pressure (meridional fronts) travelling from west to east between the migratory anticyclones and often being greatly intensified by the temperature contrast between the continent and the ocean.

That the increased summer rainfall in the south of the continent is not due to one or two exceptionally wet summers in Period 2 is shown by the frequencies of summer rainfalls given in Table 2 for the region around Adelaide, the rainfalls at Adelaide, Georgetown, Kapunda, and Cape Borda having been meaned for this purpose.

On examining the summer months individually it was found, for the whole area for which the summer rainfall in the second period exceeded that in the first by 20 per cent. or more, that the increase was greatest for February, averaging about 70 per cent., moderately large for December, 30 per cent., and negligible for January.

---

* Nearly all the places given in Table 1 do show a rise from Period 1 to Period 2 in mean winter minima, the range being from $-0 \cdot 3$ to $+2 \cdot 4$ °F., with an average of $+0 \cdot 9°$ F.

Winter precipitation (June, July, August) has changed less markedly than that of the summer between the two 30-year periods but Figure 4 shows Period 2 to have been drier over much of the area, particularly in the interior of South Australia.
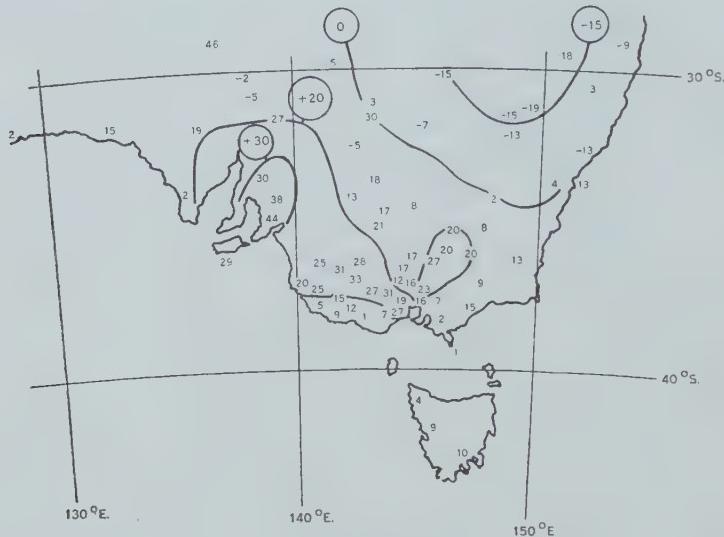
Fig. 3.—Percentage increase in summer rainfall between 1881–1910 and 1911–1940 periods.

In South Australia and western Victoria the average rainfall increases almost linearly with time during the months March to June which mark the transition from the summer minimum to the winter maximum. In Period 2 this increase has lagged behind that in Period 1 by about a fortnight. This

TABLE 2

SUMMER RAINFALL FREQUENCIES FOR REGION AROUND ADELAIDE

Rainfall (in.)	Number of Summers in	
	1881–1910	1911–1940
<1	2	1
1–1·99	15	8
2–3	7	10
>3	6	11

tendency is apparent in the graphs given by Cornish (1936) in his study of the incidence and duration of the winter rains at Adelaide.

As a result of the opposite sense of the summer and winter rainfall trends for the south of the continent, the mean annual rainfalls have changed but little

between the two periods. In South Australia the changes are negligible. In Victoria the second period has averaged a little wetter, some 7 per cent. in the central district (mean of 9 stations) but elsewhere the increases are only about 2 or 3 per cent.

V. BAROMETRIC PRESSURE CHANGES

Atmospheric pressure data show interesting differences between the means of the two 30-year periods, as will be seen from the graphs for Adelaide and Sydney shown in Figure 5. During Period 1 at both places there is a definite secondary pressure minimum in June which is absent in Period 2. The annual
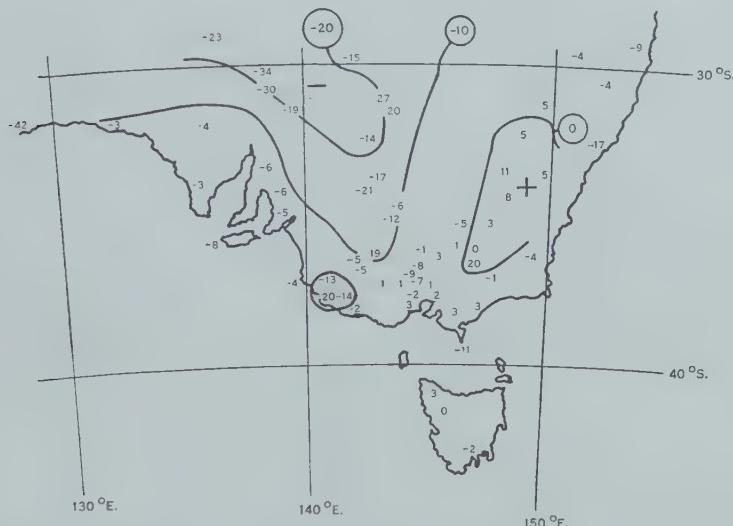
Fig. 4.—Percentage increase in winter rainfall between 1881–1910 and
1911–1940 periods.

oscillation in latitude of the high pressure belt was evidently such that its axis was appreciably to the north of Adelaide (35 °S.) and Sydney (34 °S.) in June during the first period, whereas these places mark approximately its most northerly average position in the second period.

The data for Wellington, N.Z., exhibit a similar effect but, owing to the higher latitude, the secondary minimum in winter was present in both periods. It was, however, considerably more pronounced in the first period than in the second.

This comparison of the form of the annual pressure variation has the advantage of avoiding a comparison between absolute pressure values for the two periods ; such comparisons may frequently be doubtful owing to imper-fections in the instrumental techniques of the earlier period. When, however, it is the change in pressure from one month to another that is compared between periods, then changing technique can be expected to have only a negligible influence.

Making the assumption that the difference in form between the graphs of Figure 5 was caused by displacement of the high pressure belt and not by a change in the form of its meridional profile, it was estimated that the southward displacement in the mean northerly limiting position of the high pressure axis was some 3° of latitude or 200 miles between the two periods.
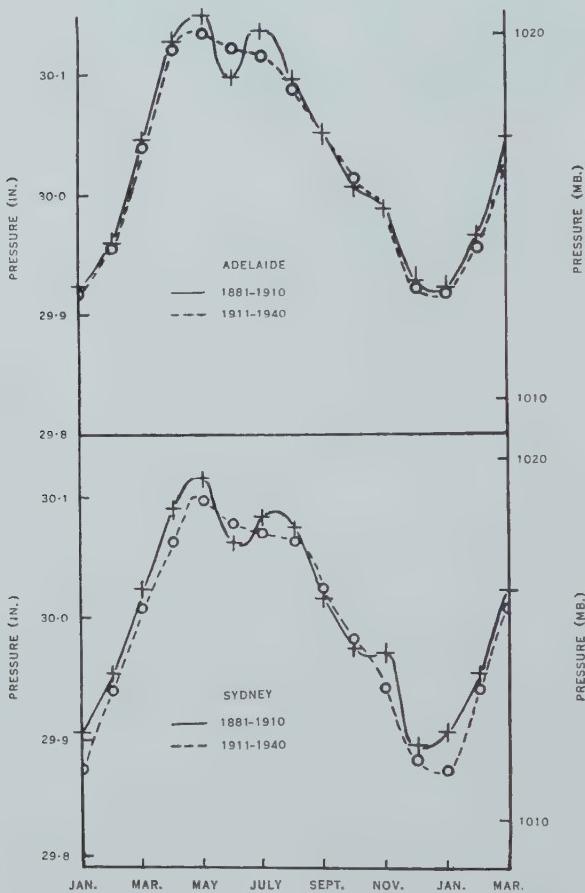
Fig. 5.—Comparison of 30-year means of annual pressure cycle at Adelaide and Sydney.

VI. DISCUSSION

As evidence has been found above for an appreciable climatic change in the Australian region contemporaneous with the notable amelioration of climate in higher latitudes of the northern hemisphere, it is desirable to discuss the present work in relation to the main findings as to the type of change in the northern hemisphere. Until recently it was far from clear whether the rising winter temperatures in north-west Europe and elsewhere were a consequence of increased activity of the zonal westerly winds or whether increased meridional transport, usually associated with a more broken and less intense zonal circulation, was

responsible.   Pettersen (1949) has, however, made use of the northern hemisphere charts of the Historical Weather Map Series of the U.S. Weather Bureau to investigate more thoroughly the changing pattern of zonal motion and meridional interchange for the North Atlantic region in the period 1899–1939.   He concluded that it was not possible to ascribe the increased temperatures in high latitudes to a speeding-up of the general westerly winds in middle and high latitudes and found the most systematic and pronounced change in the general circulation to consist in " a vast intensification of the rate of total transport between the cold arctic source and middle latitudes, and a much smaller intensification of the total exchange across subtropical latitudes ".

The increased summer rainfall in South Australia and Victoria in 1911–1940 as compared with 1881–1910 is most probably a result of increased activity in the troughs between the migratory anticyclones and it points to an increase in meridional interchange, which, together with the resultant increase in cloud amount, would account for the decreased inland maximum temperatures at this season.   If this interpretation is correct then it suggests that, in these southern latitudes, the climatic trend was of a similar basic nature to that in the North Atlantic region.   Unfortunately, it is impossible to obtain much information of occurrences in higher southern latitudes owing to lack of observations but it may be noted, however, that Willett *et al.* (1949, p. 49) find some evidence (presumably based, almost entirely, on observations in the South American sector) that during 1910–1934 the mean spring pressure at 50–60 °S. was distinctly lower than in the period 1873–1909, and this suggests some change in pattern or intensity of the westerlies.

The Australian pressure data suggest that the high pressure belt did not move so far north in Period 2 as in Period 1.   Some evidence that pressure changes at about 35 °S. in the neighbourhood of South Africa, South America, and Australia were similar in some respects is afforded by a pressure change chart given by Lysgaard (1950) who used the same two periods as in the present work. His chart for January shows that in the second period, as compared with the first, pressures at 30–35 °S. have fallen on the eastward side of each of the three land masses and risen or remained steady on the westward side, so that in all three cases there has been an increase in the mean southerly geostrophic wind component at this time of year.

The differences in 30-year mean pressure gradients between Adelaide and Sydney were examined to obtain more information on this point.   There was no appreciable change in the winter months but for the summer half-year (November to April) there was an increase in the mean southerly geostrophic wind component from $0 \cdot 5$ m./sec. in the first period to $1 \cdot 0$ m./sec. in the second, which lends a little support to increased meridional interchange in the second period as compared with the first.

As the subtropical high pressure belts are stronger heat sources than the equatorial belt, it seems that increase in meridional heat transfer is compatible with the rising trend of temperatures in Indonesia which have been studied by de Boer and Euwe (1949) and Schmidt-ten Hoopen and Schmidt (1951).

VII. CHANGES SINCE 1940

The data for the decade 1941–1950 (referred to hereafter as the last decade) were examined to see whether or not the trend displayed by the earlier observations continued.

The changes in summer rainfall from Period 2 to the last decade have followed a mainly similar pattern to that of Figure 3 and, over the large area in that figure enclosed by the 20 per cent. isopleth, the rainfall of the last decade has averaged about 20 per cent. greater than in Period 2. Around Adelaide (4 stations) the mean increase was 19 per cent., while in western Victoria it was 22 per cent. (12 stations). Over this region, therefore, the last decade had an average summer rainfall nearly 50 per cent. greater than in the period 1881–1910.

The mean summer maximum temperatures at the places in Table 1 were mainly lower in the last decade than in Period 2, with the exception of the more easterly stations Bathurst, Goulburn, and Cooma, which showed rises of $0 \cdot 3$, $1 \cdot 5$, and $1 \cdot 3$ °F. respectively. The average change at the other 11 stations was a fall of $0 \cdot 74$ °F., so the general indication is that the trend shown by the earlier data continued into the last decade over much of the area.

The mean southerly geostrophic wind component (November to April) between Adelaide and Sydney was greater in the last decade as compared with Period 2 by $0 \cdot 1$ m./sec., a change in the same sense as that from Period 1 to Period 2.

As all these differences are of a similar character to that between Periods 1 and 2 it seems that the main trend up to 1940 was continued in the decade 1941–1950.

VIII. ACKNOWLEDGMENT

Acknowledgment is made to the Director, Commonwealth Meteorological Branch, for making available climatological data used in this work.

IX. REFERENCES

AHLMANN, H. W. (1949).—*Geogr. J.* **62** : 165.

DE BOER, H. J., and EUWE, W. (1949).—Verh. Magn. Met. Obs. Batavia No. 35.

BROOKS, C. E. P. (1949).—" Climate through the Ages." 2nd Ed. (Ernest Benn : London.)

CORNISH, E. A. (1936).—*Quart. J. R. Met. Soc.* **62** : 481.

HANZLIK, S. (1931).—*Beitr. Geophys.* **29** : 138.

LYSGAARD, L. (1950).—Centenary Proc. Roy. Met. Soc., p. 206.

PETTERSEN, S. (1949).—*Geogr. Ann. Stockh.* **1949** (H. 1-2) : 212.

SCHMIDT-TEN HOOPEN, K. J., and SCHMIDT, F. H. (1951).—Verh. Kementerian Perhubungan Djawatan Meteorologi dan Geofisik, Jakarta No. 41.

WILLETT, H. C., *et al.* (1949).—Final Report of the Weather Bureau—M.I.T. Extended Forecasting Project for the Fiscal Year July 1, 1948-June 30, 1949, Cambridge, Mass.

THE SOUTHERN AURORAL ZONE AS DEFINED BY THE POSITION OF HOMOGENEOUS ARCS

By F. Jacka*

Summary

Observations on homogeneous arcs (H.A.) made at Macquarie Island (geomagnetic coordinates 60·7 °S., 243·1 °E.) are examined. The latitude of H.A. is found to be significantly dependent on the geomagnetic planetary disturbance index K_p and the time of day. For given values of K_p and time the probability distribution of latitude of H.A. has a sharply defined mode about 0·5° north of the mean.

I. Introduction

The main features of the frequency distribution of geographic position of auroras in the northern hemisphere are well known from the work of Fritz (1881) and Vestine (1944). The first comprehensive study of this distribution in the southern hemisphere was made by White and Geddes (1939) who found the zone of maximum frequency of occurrence to be roughly circular with a radius 18° about a point (75 °S., 129 °E.) near the magnetic axis pole. Vestine and Snyder (1945), from an analysis of the disturbance daily variation (S_D) of the Earth's magnetic field and comparison with northern hemisphere observations, estimated the positions of the zone of maximum frequency and other isochasms (lines of equal frequency of occurrence of auroras). Their estimates show only very rough agreement with the observed frequencies of occurrence of auroras up to that time (cf. Fig. 1). In all of these studies the data were taken from various periods in the sunspot cycle, and the auroral observations referred to auroras of all forms.

In the present paper attention is confined to just one type of aurora, viz. homogeneous arcs (H.A.) (see Atlas of Auroral Forms 1930), and the latitude of the southern auroral zone as defined by the most frequent position of H.A. on the 243 °E. geomagnetic meridian is determined by an examination of the records of 211 H.A. observed by Parsons and Fenton (1953) at the Australian National Antarctic Research Expedition Station at Macquarie Island during the period May 1950 to April 1951. The geographic coordinates of Macquarie Island are 54° 29′ S., 158° 58′ E., the geomagnetic coordinates being 60·7° S., 243·1° E. The data used are the observers' estimates of the elevation above the southern horizon of the apex of the lower border of the H.A., the fraction in eighths of the sky which was cloud covered, the hour (G.M.T.) during which the observation was made, and the value to the nearest integer of K_p (geomagnetic planetary disturbance index) associated with the 3-hr. period including the hour of observation.

* Antarctic Division, Department of External Affairs, Melbourne.

The observations on H.A. were selected for the following reasons :

(1) The orientation and simple shape of H.A. permit of fairly accurate visual estimation of the elevation of the apex of the lower border which was in all cases approximately in the magnetic meridian plane.   The observers consider that errors of more than 5° in these estimates are rare.
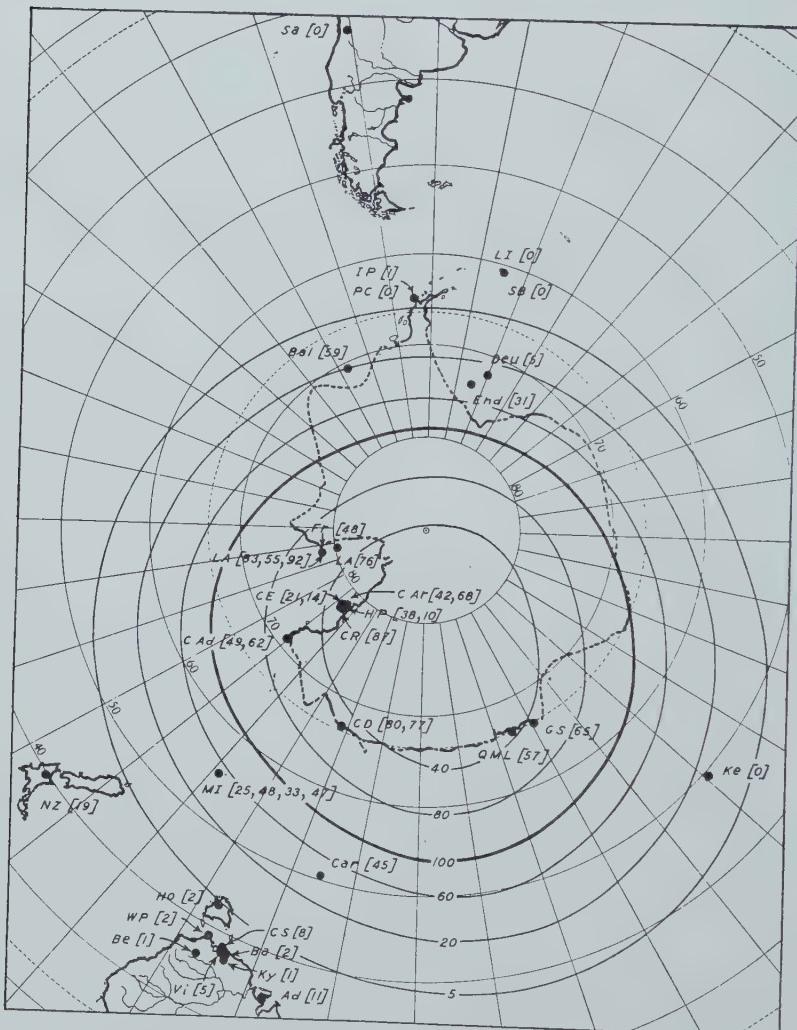
Fig. 1.—Diagram by Vestine and Snyder (1945) showing estimated percentage frequency of days with occurrence of aurora on clear dark nights at high latitudes, southern hemisphere. Figures in square brackets are observed values.

(2) H.A. appear to be due to a quiet discharge (as distinct from a constricted discharge which may remain fixed in the atmosphere for long periods). This has also been suggested by Alfvén (1950) who considers H.A. to define the position of the " A curve " of his theory.

The latitude departures $L°$, measured positive to south (see Fig. 2), of the H.A. from Macquarie Island were calculated from the relation

$$\cos(L+a) = (1+h/R)^{-1} \cos a, \quad \cdots\cdots\cdots\cdots\cdots \text{ (1)}$$
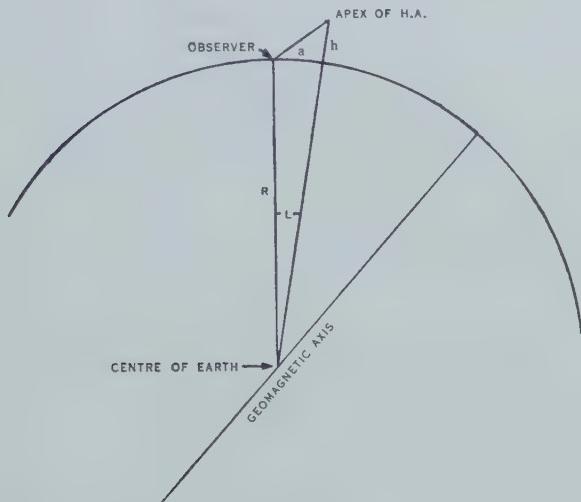
Fig. 2.—L is the latitude departure of the H.A. from the observer at Macquarie Island.

where a=elevation of H.A., R=radius of the Earth ($6 \cdot 36 \times 10^3$ km.), h=height of the lower border of H.A., taken as 105 km. (near the most frequent value found in the northern hemisphere (cf. Harang 1951)).
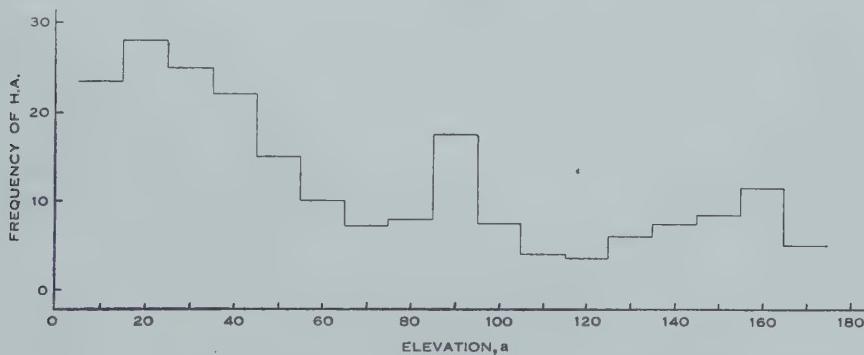
Fig. 3.—Observed frequency of H.A. in 10° intervals of elevation measured from the southern horizon.

The observed frequency of H.A. in 10° intervals of elevation is shown as a function of elevation in Figure 3. In Figure 4 (a) the relative frequency of H.A. per degree of latitude is shown as a function of L. This curve clearly shows two maxima, one at $L \simeq 1°$ and one at $L=0°$. The possibility has been considered that the decrease in relative frequency south of $L=1°$ may be due to the effect of cloud obscuring the sky at low elevations. The unimportance of
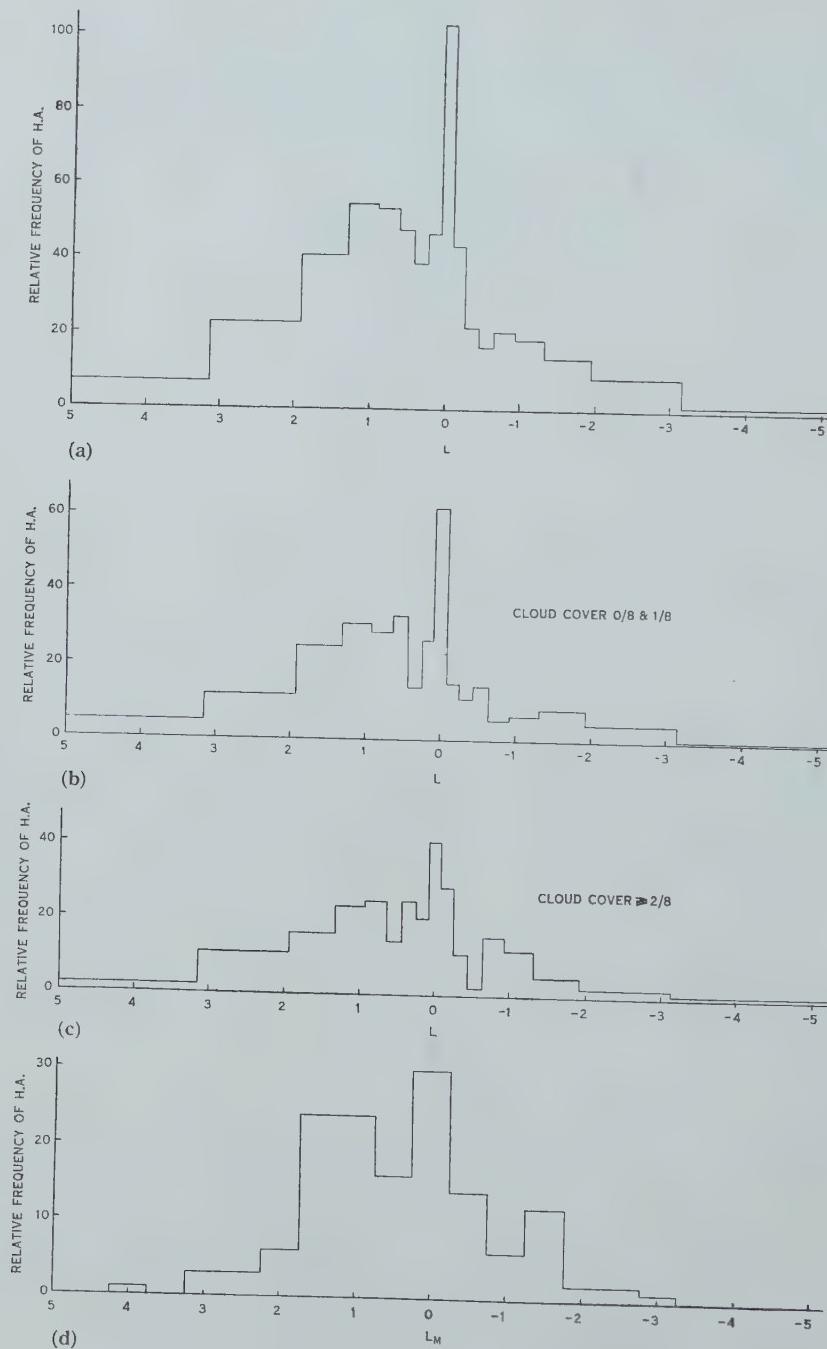
Fig. 4.—Relative frequency of H.A. per degree of latitude departure L.
(a) Derived from all observations; (b) derived from observations made at
times when the cloud cover was 0/8 or 1/8; (c) derived from observations
made at times when the cloud cover was $\geqslant 2/8$; (d) calculated from observed
values of K_p and t using equation (5).

cloud effects is, however, clearly demonstrated by the similarity of Figures 4 (*b*) and 4 (*c*), which were derived from observations made at times when the cloud cover was 0/8 or 1/8 and ≥2/8 respectively.

II. LATITUDE OF H.A. AND GEOMAGNETIC DISTURBANCE

Several workers in the northern hemisphere have shown that the auroral zone expands with increase in geomagnetic activity. Among them Rostad (1935) found a roughly linear relation between auroral zone radius and magnitude of the perturbing magnetic vector at Potsdam and Rude Skov. Nagata (1950), from an examination of magnetic records of a number of stations at different times during one magnetic storm, found the mean radius of the auroral zone, as defined by the line of maximum geomagnetic north component of the D_{st} field ΔX_m, to increase during the development of the storm and to be related to the mean equatorial ΔX_m. This relation was interpreted using Störmer's auroral theory and found consistent with the view that increase in auroral zone radius is due to modification of the Earth's field associated with the development of an equatorial ring current of radius about 20 Earth radii. The solar particle stream was assumed to consist of Ca^+ ions moving with velocity 1000 km./sec., the particle trajectories being determined by the magnetic field of the Earth.

According to Martyn's (1951*a*, 1951*b*) theory one would expect the radius of the outer edge of the auroral zone to increase with increase in kinetic energy density of the solar particle stream. Also Alfvén's (1950) theory indicates an increase in mean radius of his " *A* curve " with increase in velocity of the stream.

The geomagnetic planetary disturbance index K_p was devised to measure the intensity of the solar particle stream by its geomagnetic effects (cf. Bartels and Veldkamp 1949 ; Bartels 1949). For this reason it was considered desirable to examine the relation between latitude of H.A. and K_p.

Figure 5 shows the mean latitude of H.A. associated with each value of K_p, the straight line

$$L_E = 2 \cdot 95 - 0 \cdot 605 K_p \quad \dots\dots\dots\dots\dots\dots (2)$$

through the points being fitted to the raw data by the least squares method. One of the assumptions on which the validity of the least squares method is based is that successive errors are distributed independently of one another. When this condition is violated the estimates of the regression coefficients, though unbiased, need not have least variance and the t and F tests generally used for making confidence statements are not valid.

In the present case the observations were ordered chronologically and where several observations occurred during the same hour these were shuffled so that their ordering was random. The deviations from the regression equation (2) then show positive serial correlation significant at the 1 per cent. level. This is indicated by the statistic $d = \Sigma(\Delta Z)^2/\Sigma Z^2 = 1 \cdot 43$ being less than $d_L(1\%) > 1 \cdot 52$ as given by Durbin and Watson (1951). (Z is deviation from regression, ΔZ's are first differences of Z). Consequently the usual test of significance of the correlation coefficient r_{LKp}, according to which in this case the value found is significant at the 1 per cent. level, is not valid. The deviations Z from regression

show, on the average, a systematic non-linear variation with time of day as shown in Figure 6. A regression equation linear in K_p and quadratic in time t was therefore fitted, the equation found being

$$L_E = 2 \cdot 72 - 0 \cdot 72 K_p + 0 \cdot 06(t - 12 \cdot 5) + 0 \cdot 104(t - 12 \cdot 5)^2. \quad \ldots (3)$$

The value used for t is the mid-point of the hour (G.M.T.) of observation. In this case, too, the deviations from regression show positive serial correlation significant at the 1 per cent. level; $d = 1 \cdot 52 < d_L(1\%) \simeq 1 \cdot 60$. An examination of the values of ΔZ shows that in a number of places several small values follow one another, these being associated mainly with the occurrence of several observations during the same hour.
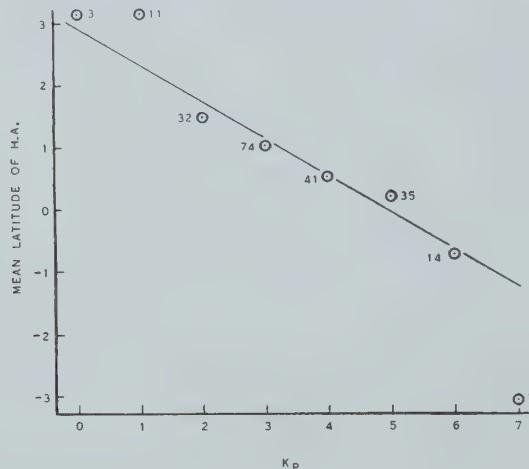
Fig. 5.—Dependence of latitude of H.A. on K_p. (The figures near the points indicate the number of observations used in computing the point.)

For this reason a regression equation linear in K_p and quadratic in t was fitted, taking as dependent variable the mean latitude L' of all H.A. observed during the same hour. Cases where several observations were made during the same hour are mainly those where several H.A. occurred simultaneously. The observations were again ordered chronologically. The regression equation obtained is

$$L'_E = 2 \cdot 98 - 0 \cdot 79 K_p + 0 \cdot 06(t - 12 \cdot 5) + 0 \cdot 11(t - 12 \cdot 5)^2. \quad \ldots (4a)$$

The deviations from regression in this case show no serial correlation significant at the 5 per cent. level; $d = 1 \cdot 781 > d_U \simeq 1 \cdot 780$. Equation (4a) may be rewritten in the form

$$L'_E = 2 \cdot 97 - (0 \cdot 79 \pm 0 \cdot 20) K_p + (0 \cdot 11 \pm 0 \cdot 05)[t - (12 \cdot 2 \pm 0 \cdot 8)]^2,$$

$$\ldots \ldots \ldots \ldots (4)$$

in which the limits shown are the 95 per cent. fiducial limits.

All of the above statistical tests involve the assumption that the errors (departures from regression in the population) are normally distributed with

constant variance.   An examination of the frequency distribution of $L-L'_E$
(Fig. 7) suggests that this is not so in the present case ; however, this is not
likely to affect the conclusions here.

The quadratic form of equation (4) was chosen purely on the grounds of
simplicity to fit the data which cover the range $K_p=0$ to 7, $t=7\cdot5$ hr. to $17\cdot5$ hr.
G.M.T. ; it is not inconsistent with the view that the auroral zone is a closed
curve, as is required by Alfvén's and Martyn's theories.

From equation (4) it can be seen that for any K_p, L'_E has a minimum value
at $t_{min}=12\cdot2\pm0\cdot8$ hr. G.M.T. which is not significantly different from mean
magnetic midnight ($12\cdot4$ hr. G.M.T.) as defined by McNish (1936).   This result
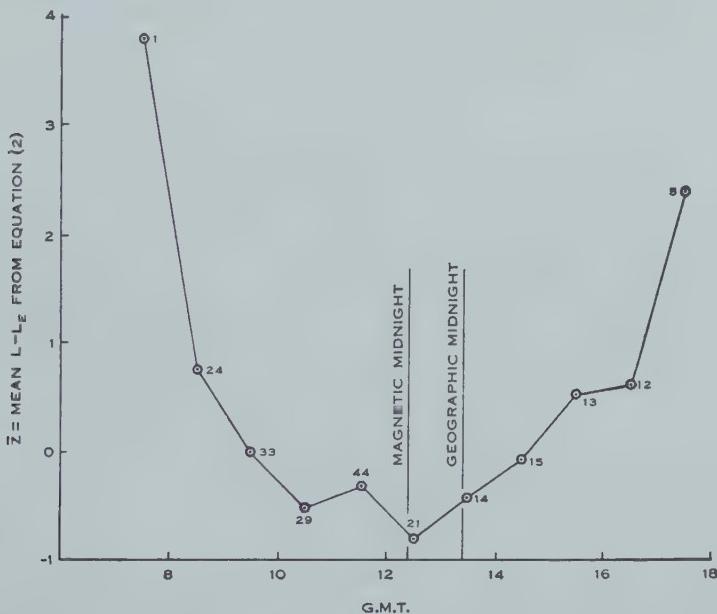
Fig. 6.—Nocturnal variation of deviation of observed L from equation (2).
(The figures near the points indicate the number of values used in com-
puting the point.)

is in fair agreement with observations on arcs in the northern hemisphere
summarized by Alfvén (1950).   According to Alfvén's theory, however, the
minimum L'_E should occur 6 hr. before magnetic midnight.   Martyn's theory,
on the other hand, indicates no dependence of L'_E on time of day.

In Figure 7 is shown the frequency distribution of deviations $L-L'_E$ of the
observed latitudes of H.A. from the regression equation (4).   The smooth curve
estimates the probability distribution of latitude of H.A. at constant K_p and t ;
but it definitely underestimates the probability at each end of the curve, especially
for $L>L'_E$.   This is caused by the limitation in range of L over which H.A. may
be observed from a single station.

At this stage it may be pointed out that when several H.A. occur simul-
taneously their latitudes are not independently distributed.   This is shown

by the fact that the deviations of observed L from regression equation (3) show significant positive serial correlation while the deviations of observed L' from regression equation (4) do not.

Figure 7 shows that the auroral zone is extremely sharply defined especially on its northern edge where the probability of occurrence of an H.A. decreases from its maximum to half the maximum over a range of less than $1\frac{1}{2}°$ of latitude. This result is of course not inconsistent with any of the auroral theories proposed by Störmer, Alfvén, or Martyn but the general shape of the curve (Fig. 7) is relevant to considerations of the accuracy of Alfvén's and Martyn's theories.

According to Martyn's theory the auroral zone is an annular ring about 6° wide. On the outside of this ring positive ions impinge on the upper atmosphere while on the inside edge electrons enter. If now we assume that both positive ions and electrons produce H.A. we would expect the distribution of L at constant
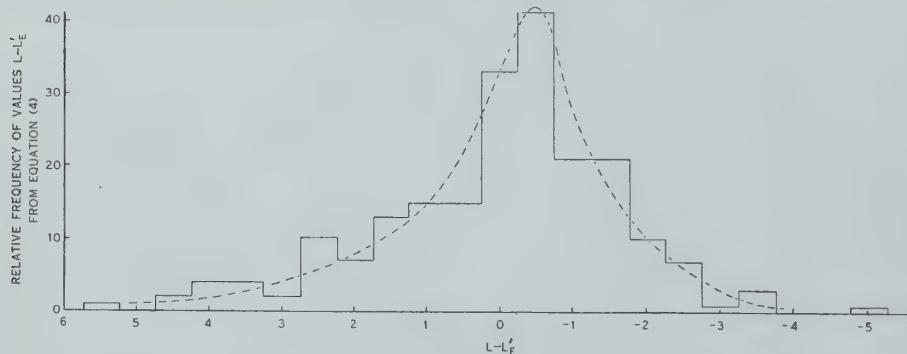
Fig. 7.—Frequency distribution of deviations $L-L'_E$ of observed latitudes of H.A. from equation (4). The dotted curve estimates the probability distribution of latitude of H.A. at constant K_p and time of day.

K_p and t to be rectangular with a width of about 6° or bimodal with the modes separated about 6°. Figure 7 is not consistent with either case. The presence of the $H\beta$ line in the spectrum of H.A. (see, for example, Dahlstrom and Hunten 1951) demonstrates that protons do occur in the streams of particles producing this phenomenon. If, then, we assume that electrons do not produce H.A. we may identify the line of maximum frequency of occurrence of H.A. with the outer edge of Martyn's auroral zone.

According to Alfvén's theory the auroral zone has a well-defined outer edge only, the day side being produced by positive ions, the night side by electrons. The results of this work, then, are consistent with Alfvén's theory only if electron streams do produce H.A.

Vestine (1944) estimated the probability distribution of latitude of the aurora in the northern hemisphere. His result is shown in Figure 8. The extreme difference between this curve and Figure 7 is largely explained by the following features of the data used by Vestine :

(1) Since the ordinate represents the probability that the aurora will be within " seeing range " from a given point the latitude resolution is about 10–12°.

(2) The curve is derived from observations associated with various values of K_p and t.

(3) The observations used refer to auroras of all forms while the probability distribution may be different for different forms.   This is certainly the case at Macquarie Island where pulsating and flaming auroras occur almost invariably north of all H.A. visible.
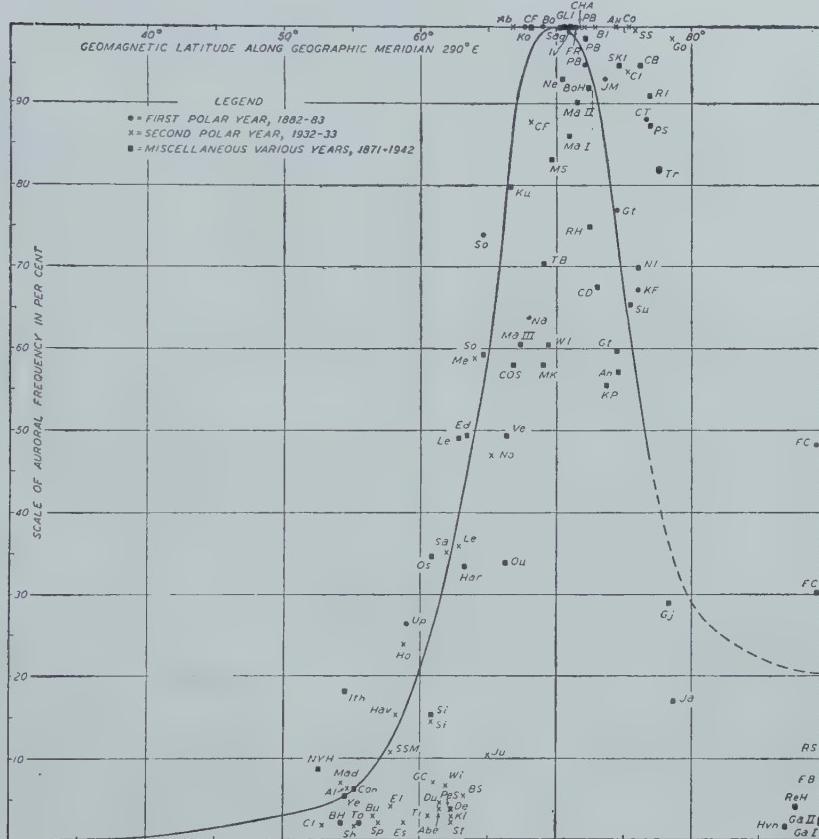
Fig. 8.—Estimated average auroral percentage frequency on clear dark nights, meridian 290 °E. ; adjusted to circular zone of maximum auroral display in geomagnetic latitude 69 °N. (after Vestine 1944).

(4) Vestine assumes the mean auroral zone (Alfvén's " M curve ") to be circular with its centre on the magnetic axis pole.   Owing to the local irregularities of the Earth's magnetic field this is unlikely to be an accurate assumption.

From an examination of the distribution given in Figure 7 it can be seen that the most probable latitude L_M of an H.A. is about $0·5°$ less than the expected value L'_E given by equation (4).   The equation

$$L_M = 2·47 - 0·79K_p + 0·11(t - 12·2)^2 \quad \cdots\cdots\cdots\cdots (5)$$

then predicts the most probable position of an H.A. given that an H.A. occurs. The frequency distribution of L_M (calculated from values of K_p and t) is given in Figure 4 (d). This clearly shows the main features of the observed distribution (Fig. 4 (a)).

Summarizing then : the expected value of the mean latitude L' of all H.A. visible from Macquarie Island at the one time is given by equation (4), while the latitude of the southern auroral zone as defined by the line of most frequent occurrence of H.A. is given on the 243 °E. geomagnetic meridian by equation (5).

III. ACKNOWLEDGMENTS

The author wishes to express thanks to Mrs. U. Brent for her assistance with numerical computations, to Mr. N. R. Parsons and Mr. K. B. Fenton for discussions on their observations, and to Dr. D. F. Martyn who read the manuscript and offered valuable comments on it.

IV. REFERENCES

ALFVÉN, H. (1950).—" Cosmical Electrodynamics," Ch. 6. (Oxford Univ. Press.)

ATLAS OF AURORAL FORMS (1930).—(Int. Un. Geod. : Oslo.)

BARTELS, J. (1949).—Geomagnetic indices, K and C, 1948. Appendix B : The standardised index K_s and the Planetary index K_p. Association of Terrestrial Magnetism and Geophysics (Int. Un. Geod.) Bulletin No. 12b.

BARTELS, J., and VELDKAMP, J. (1949).—International data on magnetic disturbances. First quarter, 1949. *J. Geophys. Res.* **54** : 295-9.

DAHLSTROM, C. E., and HUNTEN, D. M. (1951).—O_2^+ and H in the auroral spectrum. *Phys. Rev.* **84** : 378-9.

DURBIN, J., and WATSON, G. S. (1951).—Testing for serial correlation in least squares regression. II. *Biometrika* **38** : 157-78.

FRITZ, H. (1881).—" Das Polarlicht." (F. A. Brockhaus : Leipzig.)

HARANG, L. (1951).—" The Aurorae." Ch. 2. (Chapman and Hall : London.)

McNISH, A. G. (1936).—Geomagnetic coordinates for the entire Earth. *Terr. Magn. Atmos. Elect.* **41** : 37-43.

MARTYN, D. F. (1951a).—Magnetic storms and aurorae. Solar particles at the Earth. Mixed Commission on Ionosphere, Int. Coun. Sci. Un., Proc. 2nd Meeting, Brussels, 1950, pp. 49-58.

MARTYN, D. F. (1951b).—The theory of magnetic storms and auroras. *Nature* **167** : 92-4.

NAGATA, T. (1950).—Development of a magnetic storm : the southward shifting of the auroral zone. *J. Geophys. Res.* **55** : 127-41.

PARSONS, N. R., and FENTON, K. B. (1953).—Observations of the Aurora Australis, Macquarie Island, May 1950-April 1951. Aust. Nat. Antarctic Res. Expedition Interim Report No. 5.

RØSTAD, A. (1935).—Beziehungen der Nordlichterscheinungen zu den weltweiten magnetischen Störungen. *Geofys. Publ.* **10** (10) : 1-10.

VESTINE, E. H. (1944).—The geographic incidence of aurora and magnetic disturbance, Northern Hemisphere. *Terr. Magn. Atmos. Elect.* **49** : 77-102.

VESTINE, E. H., and SNYDER, E. J. (1945).—The geographic incidence of aurora and magnetic disturbance, Southern Hemisphere. *Terr. Magn. Atmos. Elect.* **50** : 105-24.

WHITE, F. W. G., and GEDDES, M. (1939).—The Antarctic zone of maximum auroral frequency. *Terr. Magn. Atmos. Elect.* **44** : 367-77.

RADAR OBSERVATIONS OF RAIN AT SYDNEY, NEW SOUTH WALES

By G. A. Day*

[*Manuscript received December* 4, 1952]

Summary

The radar echoes received from rain during systematic observations carried out from January 1950 to January 1951 at the Radiophysics Laboratory, Sydney, N.S.W., are illustrated and described. Examples of echoes from melting and upper bands, non-freezing and frontal showers, thunderstorms, and echo masses are shown by vertical cross sections of the atmosphere through the point of observation.

Correlation between the rain echo and the prevailing weather situation shows that the melting-band rain, which over this period was responsible for most of the rainfall, is usually associated with low pressure, whereas non-freezing showers are shown to be more frequently associated with high pressure.

I. Introduction

It has been well established that various forms of precipitation give rise to radar echoes. Ryde (1947) and others have given a theoretical derivation of the intensity of signal received from the scattering of radio energy by raindrops, ice crystals, and snowflakes. It is shown that the reflected radio power received at the radar set is a function of the constants of the equipment and the range to the raindrops and is proportional to the summation $\Sigma N D^6$, where N is the number of drops, of diameter D, per unit volume, the summation being taken over the whole range of the drop sizes.

The sixth-power law which determines the echo intensity has the effect that, with the wavelength and equipment used, radar echoes are only received from actual precipitation, and that no echoes are received from masses of cloud droplets even at minimum range. It is this feature which enables radar techniques to be used with advantage in the study of rain physics.

Although illustrations and explanations of the most noticeable characteristics of rain echoes have already been given by Jones (1950) and others, a number of important features seen during systematic observations in this locality do not appear to have received attention. The purpose of this paper is therefore to illustrate, classify, and describe the various rain echoes received during observations carried out from January 1950 to January 1951 at the Radiophysics Laboratory, Sydney, N.S.W., and to outline an attempt to correlate the rain echoes with the weather situation prevailing at that time.

With few exceptions, radar observations were carried out whenever precipitation appeared likely between 0900 and 1800 hours each day throughout

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

the period, rain echoes being recorded on 74 separate days.   The rainfall on these
days appears to give a fair sampling of the types of rain experienced over the
period in this locality.

II. EQUIPMENT AND WEATHER INFORMATION

The equipment consists of a search radar, type SCR717B, operating with a
wavelength of 9·1 cm.   The peak power of the radar transmitter is approxi-
mately 40 kW. and the pulse duration 1·25 μsec.   The antenna system, a
dipole-fed paraboloid, is arranged to rotate about a horizontal axis at 30 r.p.m.
in such a manner that the 10° radar beam scans a path from horizon to horizon
through the zenith.   This results in a representation on the cathode-ray display
tube of a vertical cross section through the atmosphere about the point of
observation.   This type of display has an advantage over the conventional
Range Height Tube (R.H.T.) presentation, since the range and height are
maintained in true proportion, and a better interpretation of the rain echoes
immediately above the point of observation is possible.

The sensitivity of the radar equipment is such that a precipitation rate of
0·02 in./hr., assuming the rainstorm completely fills the radar beam and has a
size distribution as given by Laws and Parsons (1943), will just be detected at
a range of 10,000 ft.   Likewise, drops all of 1 mm. diameter would just be
detected at this range if their density were 100 drops/m.^{3}.   Because of the double
divergence of radio energy when using radar techniques, the intensity of the
received echo decreases rapidly with increase in range and causes even widespread
rain to appear, on the radar display, as of limited horizontal extent centred
about the point of observation.

Facilities are provided for taking 35 mm. photographs of the cathode-ray
display tubes every second, fifth, or tenth antenna revolution, and such photo-
graphs are used to illustrate this paper.   The bottom edge of each photograph
corresponds to ground level ; the radar is located at about the centre.   Small
echoes, shown in some photographs as rising from the ground level in the form
of small arcs of circles, are from nearby buildings and should not be confused
with the rain echoes.   The range markers visible on some photographs indicate,
as usual, distance from the point of observation, which, in this type of display,
corresponds to height above site level when the beam is pointing towards the
zenith.

Automatically recorded on the 35 mm. films, but not shown in the repro-
ductions, are a vertical radar " A " scope display, chronometer, date card,
and aximuth bearing of the antenna.   During observations this azimuth bearing
or plane of rotation of the antenna is aligned with the direction of movement of
the cloud mass from which the echoes are being received, so that the build-up
and decay of these echoes can be seen to advantage.

The data relating to the prevailing weather situation issued by the New
South Wales Branch of the Commonwealth Meteorological Bureau consist of
the surface synoptic weather map, the weather reports, and the record of the
radiosonde sounding taken daily at the Royal Australian Air Force station,
Rathmines, N.S.W.   At the Laboratory, rainfall and rain intensity measurements

are made and recorded, with personal observations on the weather situation generally. In addition, information on cloud structure and formations and reports on turbulence have been received from aircraft flying in the area.

III. TYPES OF RAIN ECHO

The radar echoes received from precipitation appear to fall naturally into three distinct groups. The first group comprises the echoes which are received from widespread steady rain ; the predominant feature of these is a band of high intensity situated just below the 0 °C. isotherm. A number of investigators (Ryde 1947 ; Austin and Bemis 1950 ; Hooper and Kippax 1950) have given explanations for this intensification, and agree that the change in phase of the precipitation from the solid to the liquid state and its change in vertical velocity are contributing factors to the phenomenon. In the literature the phenomenon has been called the " bright band " or " melting band " ; the latter term is used in this paper and is considered preferable as it distinguishes it from other bands of increased intensity which sometimes appear with this type of rain echo.

The second group of rain echoes consists of those which develop as individual column-like structures. These are without doubt from rain showers and they are therefore broadly classified as " shower echoes ". There are at least three types of shower from which radar echoes are received ; firstly, that in which the precipitation apparently consists wholly of water drops, as will be seen later, and in which the icing stage has played no part ; secondly, that accompanying cold frontal conditions ; and finally, the thunderstorms. The echoes received from these showers are called non-freezing shower echoes, cold-front echoes, and thunderstorm echoes respectively.

The third group consists of rain echoes which are devoid of both the column-like structure of the shower echoes and the intense band of the melting-band echoes. They appear in cross section as a mass or layer of echo extending from the ground to a height usually determined by either an inversion layer or the 0 °C. isotherm. These have been called echo masses, and can perhaps best be visualized by reference to appropriate illustrations. Detailed descriptions and illustrations of all three groups of rain echoes are given below.

(a) Melting- and Upper-Band Echoes

The melting-band echoes comprise all those in which a band of increased intensity appears just below the 0 °C. isotherm, as shown in Plate 1. Observations and aircraft weather reports suggest that the rain which produces this type of echo is usually widespread. The radar display confirms these observations and shows that the vertical cross section of the rain echo is free from marked column-like structures that would suggest strong local convection currents or turbulence.

Usually the first indications of a band forming are faint echoes above the 0 °C. isotherm, falling towards the ground. At about, or just below, the 0 °C. level, the echoes develop the increased intensity and band structure which is the characteristic feature of this formation, and which has been identified as the melting band. Following the development of this band, echoes appear falling

from the underside in the form of diffuse patches, which eventually reach ground level. The band persists, perhaps intermittently, for the duration of the rain period, which may be several days.

The observations throughout the year agree with those of Hooper and Kippax (1950) that the intense band is invariably situated just below the 0 °C. isotherm, confirming that this is the region of transition from the solid to the liquid state. The 0 °C. isotherm in this area varied throughout the year from 5000 to 16,000 ft.

It seems clear that in this melting-band rain, the precipitation originates as ice at some height above the 0 °C. isotherm. That is the mechanism of rain formation visualized by Bergeron (1933), who has stated that in his opinion almost every real raindrop and all snowflakes originate around an ice crystal.

Frequently in the presence of melting bands the echoes above the band extend to considerable heights. The upper echoes, as they are called, sometimes take the form of horizontal striations or upper bands, which can clearly be seen to develop and gradually fall towards the melting band. A detailed description of the formation of these upper bands has been given by Bowen (1951). Plate 2 shows the formation of an upper band at a height of 8000 ft. above the melting band.

Occasionally two or more upper bands can be seen simultaneously, or the upper echoes may take various other forms such as sloping layers, diffuse patches or simply a random distribution of echoes. Plate 3, Figures 1–3, shows typical examples of such upper echoes. Usually these echoes appear to fall gradually towards the melting band, their distinctive outlines becoming less obvious as the melting band is approached. At times, when a particularly distinct upper band develops, the path traversed by the echoes can be clearly followed from their initial development above the 0 °C. isotherm down to site level, illustrating very clearly the intensification of the transition region.

Only exceptionally does the melting band show any appreciable signs of turbulence. On one such occasion the melting band, for several hours, was repeatedly seen to burst up into a very ragged formation, and then subside again to a normal melting-band formation. A similar effect is illustrated in Plate 4, Figure 1, which shows a local thunderstorm developed at the same time as a melting-band formation. The echoes forming part of the band were clearly seen to be transported from 11,500 ft. to a height of 25,000 ft., obliterating the melting band and producing this combined effect.

On two occasions a lowering of the melting band was observed. These coincided with the passing of a cold front over the area, and in each case the drop in height was of the order of 1000 ft., and the time taken about 30 min.

(b) Shower Echoes

The thunderstorm, cold-front, and non-freezing shower echoes can be readily identified by their characteristic cellular column construction. Observations on the build-up of the shower echoes, together with reports from aircraft observations, confirm the meteorological evidence that convection is the predominant factor in the development of these showers.

The rain intensities, rain echo intensities and drop size spectra (samples of the latter are taken from time to time by the stained filter paper technique) all vary considerably during the rapid build-up and decay of individual showers and between the different types of shower.   The thunderstorms give rise to intense radar echoes with clear-cut edges, to rain intensities often in excess of 2 in./hr., and to numerous drops exceeding 3 mm. in diameter.   At the other extreme, the smaller non-freezing showers are more often only of moderate rain intensity (0·2 in./hr.) and have a restricted drop size spectrum with a tendency for the larger drops to reach a maximum size of about 1·5 mm. diameter.

It has been well established by Bowen (1950) and Smith (1951) that non-freezing showers are a feature of this locality, and often a number of these can be seen together on the radar display tube.   Plate 4, Figure 2, shows examples of these non-freezing showers in which the echo-producing precipitation elements develop wholly below the 0 °C. isotherm, presumably due to coalescence.   This type of shower is of frequent occurrence in this locality during the warmer months of the year, and situations favourable for its development may last for several successive days.   A well-marked temperature inversion below the 0 °C. isotherm invariably limits the maximum height reached by non-freezing showers to the inversion layer, as shown in Plate 4, Figure 3, where the inversion layer at 8000 ft. is some 5000 ft. below the 0 °C. isotherm ($+\frac{1}{2}$ °C. at 8000 ft., $+3$ °C. at 9000 ft.).   Aircraft observations in this area suggest that usually the atmosphere above such inversions consists of clear air free from ice crystals or cloud layers.

On occasions, shower echoes are seen to extend up to and beyond the 0 °C. isotherm, and there appear to be two distinct phenomena involved.   In some cases an obvious brightening or intensification at the top of each shower is clearly visible, at about or just below the 0 °C. isotherm, as shown in Plate 5, Figure 1, and on these occasions the melting stage is obviously part of the mechanism of rain formation.   However, there is, generally, no visible sign of increased intensity which could suggest a transition from the ice stage and, as these showers are frequently associated with those which develop wholly below the 0 °C. isotherm, the observations tend to confirm the suggestion by Bowen (1950) and Smith (1951) that showers in which the icing stage plays no part do extend, in this area, up to and beyond the 0 °C. isotherm.

An approaching cold front, distinguished visually by the characteristic frontal cloud formation, almost invariably gives rise to showers and shower echoes.   Apart from the unmistakable column construction, these cold-front showers can be readily recognized because a number of them appear together, and the radar picture suggests a series of separate convection showers embedded in the cloud, which is usually continuous, associated with the front. Observations of the development of these cold-front shower echoes, and of the cloud formations producing them, suggest this marked convection, and, although when fully developed the echoes extend from well above the 0 °C. isotherm to the ground, no apparent melting stage is visible at any level.   Typical examples of these cold-front showers are illustrated in Plate 5, Figure 2, and these particular examples clearly show the tilting effect of the lower winds on the column-like

structures. These frontal showers have a more rapid development, extend to greater heights, and are usually of greater intensity than the non-freezing showers, and, although individual showers may build up and decay rapidly, the showers as a whole may last several hours.

The difference between the cold-front showers and the thunderstorms appears to be essentially one of size ; in addition, the latter can develop entirely independently of frontal conditions as isolated convection showers or heat storms. The immense cumulo-nimbus cloud formation which develops before the appearance of the radar echoes, the peculiar squally wind which immediately precedes the rainfall, and copious precipitation are some of the characteristics of the thunderstorm. Hail occasionally fails, and thunder and lightning are invariably present. Precipitation elements of any size first develop at some height above the 0 °C. isotherm, and the echoes received from them are seen to extend in all directions, developing into the column- or tower-like structures illustrated in Plate 5, Figure 3. Plate 6 comprises a series of photographs showing the development of this thunderstorm echo.

A fully developed thunderstorm echo may extend to a height corresponding to 35,000 ft., the maximum height of the echo being reached at about the time the first precipitation reaches the ground. On reaching this maximum height the echo commences to collapse or decay, the precipitation elements responsible for the echo falling out of the sky in a violent local storm covering a comparatively small area and lasting only a short time.

Radar echoes from lightning flashes have been seen, as reported by Ligda (1950), the echo from the path of the discharge appearing as a streak of high intensity situated or commencing near that part of the echo which first appeared.

As with the cold-front showers, there is no sign of a melting or transition stage, but this does not necessarily signify complete absence of the icing stage ; in fact, some observations suggest that the glaciation of the top of the cumulo-nimbus starts this phenomenon developing. It has been suggested that the extreme turbulence which is known to exist within this type of cloud formation may be responsible for the absence of any visible melting stage.

(c) Echo Masses

The broad layer-like formation of the echo masses can perhaps best be defined by the absence both of the column structure of the shower echo and of an intense melting band. As the name suggests, they consist of a solid mass or block of echo, limited in horizontal extent only by the sensitivity of the radar equipment. As with the non-freezing showers, there appear to be at least two distinct sets of conditions for development of echo masses, and it is by no means clear whether the two phenomena are related or brought about by entirely different mechanisms.

In Plate 7, Figure 1, the maximum height of the echo-producing elements of the echo mass is shown to be that of the temperature inversion derived from the radiosonde sounding taken about the time of the observation ($+2$ °C. at 7000 ft., $+6$ °C. at 8000 ft.). The 0 °C. isotherm is shown to be some 5500 ft. above this inversion layer. This set of conditions is clearly similar to that of

RADAR OBSERVATIONS OF RAIN

Melting-band echo from precipitation, August 1, 1950.

HEIGHT (1000 FT.)

RADAR OBSERVATIONS ON RAIN
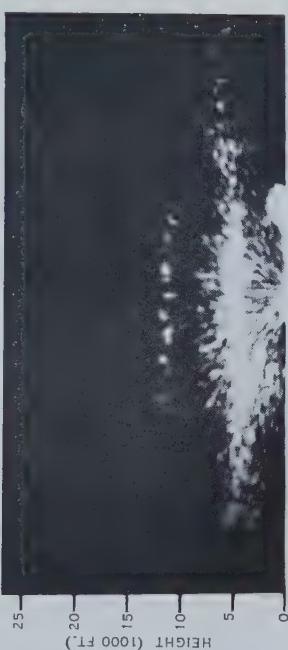
Formation of upper band, November 17, 1950, at 8000 ft. above melting band.

RADAR OBSERVATIONS OF RAIN

Fig. 1.—Multiple upper bands, July 10, 1950.

Fig. 2.—Upper echoes forming sloping layers, June 8, 1950.

Fig. 3.—Upper echoes falling into melting band, January 18, 1950,

RADAR OBSERVATIONS OF RAIN

Fig. 1.—Combined melting band and thunderstorm, December 30, 1950.
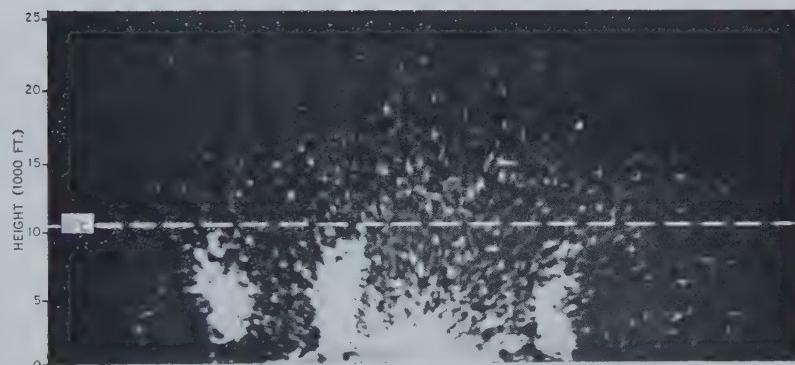
Fig. 2.—Non-freezing showers extending to 0 °C. isotherm, January 18, 1951.

Fig. 3.—Non-freezing showers limited by temperature inversion, April 14, 1950.

RADAR OBSERVATIONS OF RAIN

Fig. 1.—Shower echoes showing intensification at top, January 18, 1951.
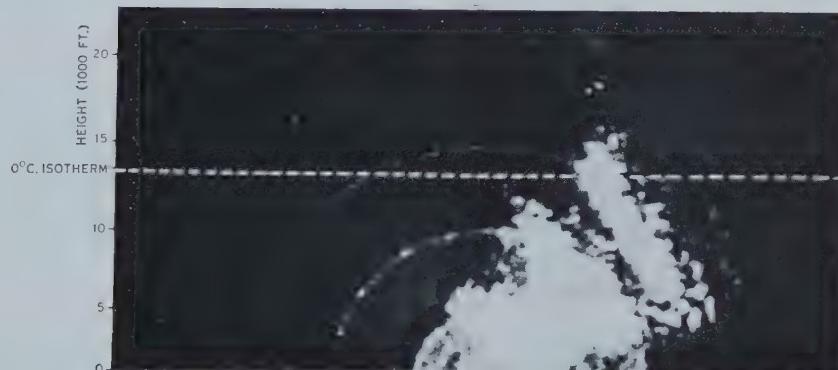
Fig. 2.—Shower echoes from cold front, March 27, 1950.

Fig. 3.—Thunderstorm echo, March 27, 1950.

RADAR OBSERVATIONS OF RAIN
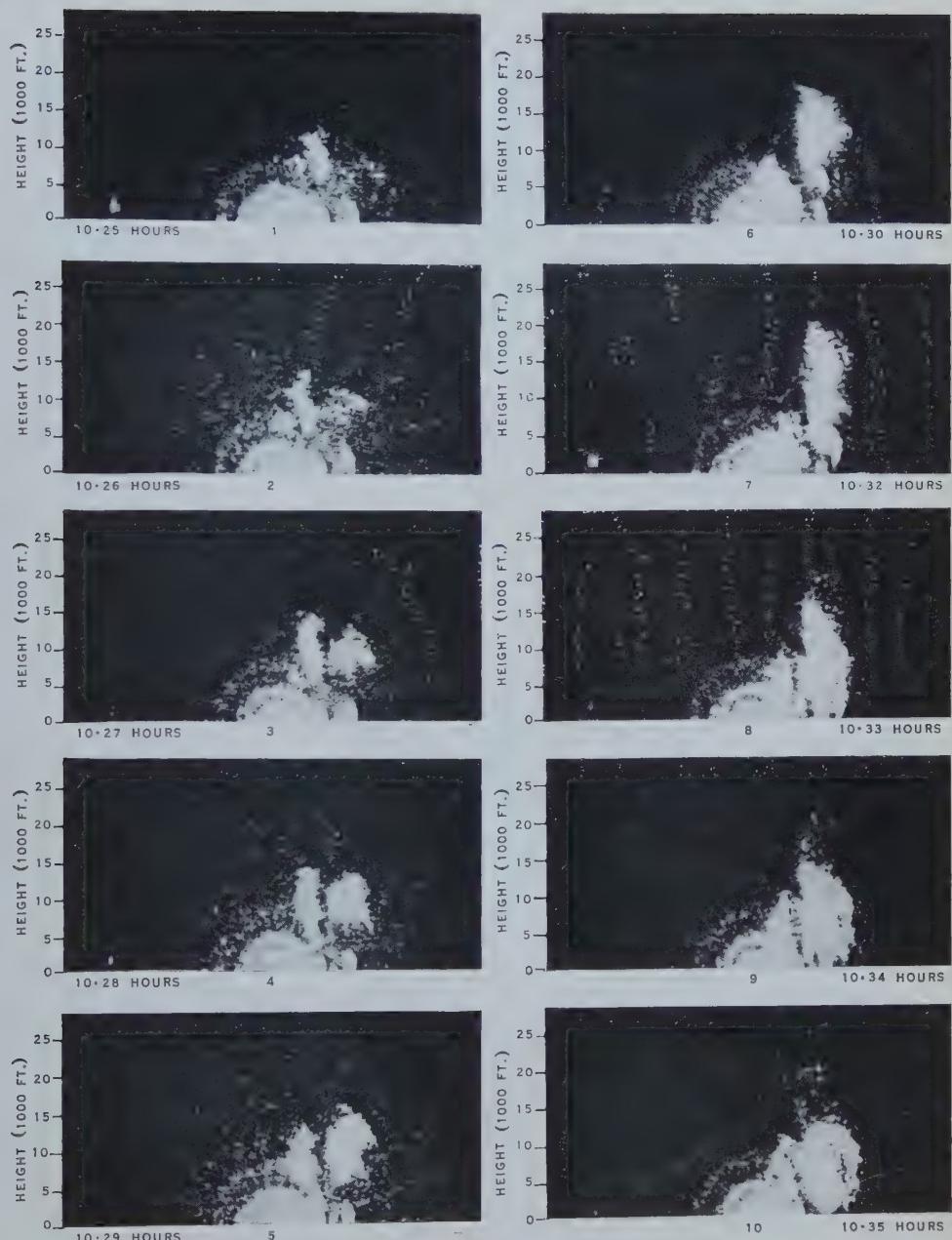
Development of thunderstorm echo.

RADAR OBSERVATIONS OF RAIN

Fig. 1.—Echo mass limited by temperature inversion, April 12, 1950.

Fig. 2.—Echo mass limited by 0 °C. isotherm, February 14, 1950.

the inversion-limited non-freezing showers, and it is likely that these masses consist wholly of water drops formed from cloud droplets by a similar coalescence process.   Drop size measurements taken when these inversion-limited echo masses prevail suggest a similar spectrum to the smaller non-freezing shower echoes, with no particularly large drops and a tendency for the largest of those present to be of uniform size.

The second type of echo mass, although it appears somewhat similar on the radar display tube, is not accompanied by an inversion, and the maximum height of the echo-producing precipitation approaches the 0 °C. isotherm.   Although at times the top of an echo mass of this type is as clearly defined as with the inversion-limited type, more frequently there is a gradual fading of the echo intensity against height, as would be expected from its range alone.   Plate 7, Figure 2, is an example of this type of echo mass.   In exceptional cases patches of echoes can be seen falling into these masses from well above the 0 °C. isotherm, but without any obvious indication of a transition stage.

Although the inversion-limited echo masses may persist for many hours, this is not the case with the other type, which more frequently appears associated with melting-band formations.   This suggests that a melting stage might be involved, the size or density of the precipitation elements being such that, with the sensitivity of the particular radar used, no radar echo is visible.   Probably at least two distinct processes are involved, and further evidence is necessary for a complete understanding of the phenomenon.

IV. Correlation between Type of Rain Echo and Weather Situation

It seemed likely that the type of rain echo observed should depend primarily on the vertical stability of the air.   Observation days were first classified into two groups according to whether the rain echoes seen indicated the presence of ice crystals or not.

(i) " Bergeron rain "—days when a melting band was present.

(ii) Purely non-freezing rain—days when the echoes (showers or masses) did not extend up to the freezing level, or where they reached close to it, but without any intensification such as would be expected if ice particles were melting just below it.

All other observation days were classified as complex.   These included days when showers or thunderstorms occurred, giving echoes which reached above the freezing level, but without any band formation near it, and days on which melting bands were observed some of the time and not at others.

Some difficulty was met in finding a criterion for the vertical stability of the air.   The radiosonde data from Rathmines did not seem adequate, partly because of the displacement in space (90 miles) and in time (up to 9 hr.), and partly because of the difficulty of radiosonde observations during bad weather. Aircraft reports of turbulence were found to be too sparse and inconsistent. In the absence of direct observational data, it was necessary to fall back on synoptic criteria.   In the neighbourhood of Sydney the association between

Table 1

RADAR ECHOES, DEPARTURE OF PRESSURE FROM MONTHLY MEAN, AND RAINFALL DATA

Date	Type of Echo*	Classi-fication†	Departure of Daily Pressure from Mean for Month (mb.)	24-Hr. Rainfall (in.)	Date	Type of Echo*	Classi-fication†	Departure of Daily Pressure from Mean for Month (mb.)	24-Hr. Rainfall (in.)
Jan.					July				
18	B	B	—4·1	4·48	13	NFS	NF	+4·0	0·36
Feb.					14	S, B	C	+7·3	
4	B	B	+0·6	1·12	17	NFS	NF	+3·6	0·17
6	B	B	—3·5	1·60	18	S, B	C	+1·8	
7	B, M, CFS	C	—7·8		24	NFS	NF	+7·9	0·29
					25	NFS	NF	+7·6	0·33
8	B	B	—6·8	0·12	26	NFS	NF	+4·7	0·23
14	M	NF	+1·5	0	27	NFS	NF	+1·1	0·64
15	B, S	C	—3·2		28	S, B	C	+3·6	
18	B	B	—3·9	0·26	Aug.				
Mar.					1	B	B	—10·6	0·25
3	B	B	—5·9	0·53	10	B	B	+7·8	0
6	B, CFS	C	—3·9		16	NFS	NF	+8·8	0·14
7	B	B	+2·1	0·88	21	B, S	C	—2·9	
14	NFS	NF	+8·8	0·12	23	B, S	C	+6·2	
27	T	C	+5·6		28	B, S	C	+4·0	
28	B	B	+2·9	0·06	Sept.				
30	B	B	—1·6	0·10	11	B	B	+6·3	1·71
31	B	B	—0·2	0·02	12	NFS	NF	+1·0	0
Apr.					13	NFS	NF	+3·0	0·72
3	B	B	—6·1	0·29	14	B, NFS	C	+1·7	
4	B	B	—11·7	0·09	Oct.				
12	NFS, M	NF	+5·2	0·09	6	B	B	—0·5	0·09
14	NFS, M	NF	+6·3	0·24	18	B	B	—4·0	0·88
30	B	C	+1·1		19	B	B	—13·2	0·39
May					20	CFS	C	—15·9	
9	CFS, M	C	—10·0		25	B	B	—6·3	0·31
24	B	B	—8·9	3·15	Nov.				
June					1	B	B	—5·7	0
7	B	B	—5·3	2·05	3	B, M, S	C	—9·3	
8	B	B	—7·1	1·91	10	CFS	C	—11·4	
13	NFS	NF	+7·0	0·35	14	B, S	C	—2·6	
14	B & S	C	+5·0		16	B	B	+4·4	0·60
15	B & S	C	—2·2		17	NFS	NF	+5·7	0·01
16	B & S	C	—5·1		20	S, B	C	+2·1	
19	CFS	C	+5·5		21	B	B	+2·3	0·32
20	NFS	NF	+6·7		24	B	B	—0·2	0·60
26	S, B	C	—14·5	0·17	Dec.				
July					6	NFS	NF	+0·1	0
3	B	B	—7·7	0·01	8	CFS	C	—12·0	
10	B	B	+2·9	2·17	23	B	B	+4·2	0·12
					30	B	B	—1·7	0·27

* B, Bergeron or band echoes ; NFS, showers not extending above freezing level ; CFS, cold-front showers ; M, masses not extending above freezing level ; S, showers extending above freezing level ; T, thunderstorm.

† B, Bergeron or band ; NF, non-freezing rain not extending above freezing level ; C, complex (not plotted on Figures 1 and 2).

rain and fronts is much weaker than in higher latitudes. Most of the rain results from convergence of apparently homogeneous air masses. An attempt to relate rain echo type to the air mass as determined from the trajectory of the

lower layers failed ; the properties of the onshore maritime streams which brought most of the rain varied only slightly from one rainy day to another, and no correlation could be found with echo type.   This latter result is consistent with the fact that the stability of air masses in this region is influenced more by vertical movements associated with mass convergence and divergence than by advective effects.   In view of this and of the known connection between stability and surface pressure situation, it was decided simply to classify the days of
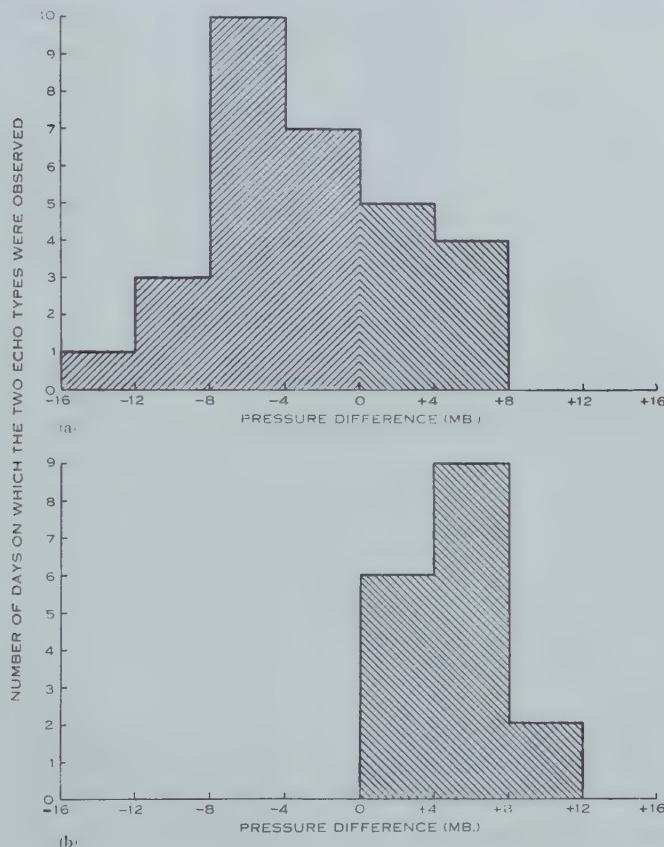
Fig. 1.—Histograms of number of days with (a) " Bergeron " and (b) non-freezing rain, grouped according to the departure of the surface pressure from the mean for the month.

observation on the basis of the pressure at Sydney.   The mean of the 0900 and 1500 hours (E.S.T.) pressures being readily available, this was used as the pressure for the day.   The mean value of these pressures was found for each month, and the departure of the day's pressure from this mean was found.

The relevant data are shown in Table 1.   For each date is given the type of echoes observed, as described in Section III above.   Then follows the classification used for the present purpose into days when the echoes showed melting bands (B), those when they were from non-freezing rain (NF) and the

undetermined or complex cases (*C*). In the fifth column is given the 24-hr.
rainfall (9 a.m. to 9 a.m.) observed at Sydney Weather Bureau for days classified
as *B* or *NF*.   For the use made of them (see Fig. 2), these rainfall measurements
may be taken as representative of the rainfall at the Radiophysics Laboratory.

The association between pressure departure and rain echo type is shown in
Figure 1, which gives histograms of days of Bergeron and non-freezing rain
grouped according to the departure of the pressure from the mean for the month.
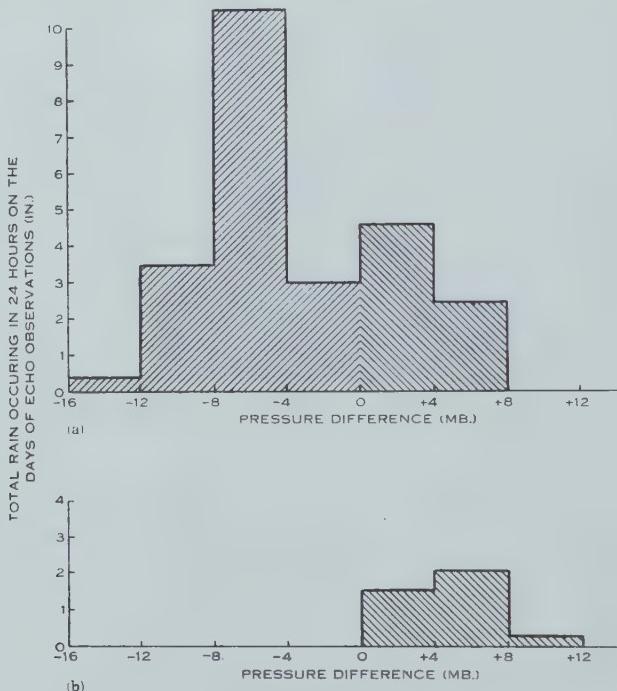
Fig. 2.—Total of 24-hr. rainfalls on days when (*a*) " Bergeron "
and (*b*) non-freezing rain fell, grouped according to the departure
of the surface pressure from the mean for the month.

Clearly, melting bands occurred mostly when the pressure was low ; non-freezing
rain occurred in this series only when the pressure was above the norm.   The
same effect is shown in Figure 2, where the total amount of rain (i.e. the sum of
the 24-hr. rainfalls) is used instead of the number of days.   Figure 2 also shows
that the days when only non-freezing rain echoes were observed accounted for
only a small proportion of the total rainfall occurring on the days of observation.
The latter amounted to 53 per cent. of the entire rainfall over the period of
observations.

V. Conclusions

From the observations described here it appears that it is possible to classify
rain echoes into several distinct types.   The distinction of greatest physical
interest is that between Bergeron-type rain, with the radar echo showing a

melting band just below the freezing level, and non-freezing rain where no such band occurs.

Bergeron-type rain was observed predominantly with low pressure, that is, in unstable air ; non-freezing rain with high pressure, that is, in stable air. It must be noted that the presence of ice crystals in a cloud does not imply that, in their absence, no rain would have fallen. The simplest interpretation of Figures 1 and 2 seems to be that, with pressure above normal, in stable air, cloud tends to be so limited in height that ice crystals are not formed and only the non-freezing rain process—the coalescence of liquid drops—can operate giving purely non-freezing rain. This latter process no doubt operates also in deeper clouds where ice is present, but the present study gives no hint of its relative importance. Thus, although Figure 2 would indicate that the relative contribution of purely non-freezing rain to the total rainfall was small, this might be quite misleading as regards the real importance of ice crystals in the rain régime at Sydney. So far as these present observations go, it could even be that, if somehow ice were prevented from forming at all in clouds, little difference would be noticed in the total rainfall.

VI. ACKNOWLEDGMENTS

VII. REFERENCES

AUSTIN, P. M., and BEMIS, A. C. (1950).—*J. Met.* **7** : 2.

BERGERON, T. (1933).—*P.V. Mét. Un. Géod. Géophys. Int.* Lisbon : 156.

BOWEN, E. G. (1950).—*Aust. J. Sci. Res.* A **3** : 193-213.

BOWEN, E. G. (1951).—*J. Atmos. Terr. Phys.* **1** : 3.

HOOPER, J., and KIPPAX, A. (1950).—*Quart. J. R. Met. Soc.* **76** : 328.

JONES, R. F. (1950).—*Met. Mag.* **97** : 934-6.

LAWS, J. O., and PARSONS, D. A. (1943).—*Trans. Amer. Geophys. Un.* **24** Part III : 452.

LIGDA, M. G. H. (1950).—*Bull. Amer. Met. Soc.* **31** : 279.

RYDE, J. W. (1947).—Meteorological factors in radar wave propagation, Rep. Phys. Soc. Lond. and R. Met. Soc., pp. 169-89.

SMITH, E. J. (1951).—*Quart. J. R. Met. Soc.* **77** : 771.

EXPLICIT MATRIX ELEMENTS FOR MULTIPOLE RADIATION

By P. B. Treacy*

Summary

Using a simple model, a formula is given for the radiation intensity from a multipole of any order. This formula is extended to include polarization properties of quanta. As an example, a matrix element is calculated for multipole radiation of order four.

I. Introduction

In the theory of angular correlations, it is relatively simple to calculate matrix elements for transitions which involve the emission of heavy particles. Thus, in the case of a nucleus of total angular momentum J and component m, which disintegrates into two particles of spins S_1 and S_2 and relative orbital angular momentum L, the lighter particle proceeding in a direction specified by spherical polar coordinates (θ, φ) in the usual way, the matrix element connecting transitions to a final state of spin $S = |S_1 + S_2|$ and component $m_S = m_{S1} + m_{S2}$ is simply

$$(Jm \mid H_L \mid Sm_S) = \alpha_{LS} \cdot (LSJm \mid LSm_L m_S) \cdot P_L^{m_L}(\cos \theta) e^{-im_L \varphi}/(2\pi)^{\frac{1}{2}}.$$

$$\dots\dots\dots\dots\dots\dots\dots \text{(1)}$$

Here $(LSJm \mid LSm_L m_S)$ is a transformation coefficient for vector addition, in the notation of Condon and Shortley (1935); α_{LS} is a complex quantity taking account of (unknown) orbital and spin dependence of the matrix element. The remaining term is $Y_L^{m_L*}(\theta, \varphi)$, a normalized spherical harmonic of order L and $m_L = m - m_S$, consistent with the definition of an arbitrary phase in $(LSJm \mid LSm_L m_S)$. (The symbol * denotes complex conjugate.) Contributions from different S and m_S values must be combined incoherently using equation (1).

It is the purpose of this note to show that formulae similar to (1) can be used in the calculation of matrix elements for multipole radiation, assuming as a model that a quantum is a " particle " of intrinsic spin one, and components $+1$, 0, and -1. With this assumption the intensity of a radiation process may be written down immediately, and a simple extension makes it possible to evaluate the polarization of the radiation field. A general formula for radiation intensity of a multipole has been given by Ling and Falkoff (1949) (see also Falkoff, Colladay, and Sells 1952). In their method polarization effects do not appear automatically. Fierz (1949) has calculated multipole matrix elements for most cases of interest, while the straightforward calculation of matrix elements of field vectors has been carried out to multipole order 3 by French (1951),

* Research School of Physical Sciences, Australian National University, Canberra,

A

using the method of Condon and Shortley (1935). The present model has the advantage of simplicity and of providing a formula completely explicit in its angular dependence.

II. Formulae for Multipole Radiation

Consider a radiative transition of order L from a system described by (J, m) to a lower state (J', m'). Because of the nature of the electromagnetic interaction, we may define the matrix element of the radiation field $\mathbf{E}_L$ (cf. Ling and Falkoff 1949) as

$$(Jm \mid \mathbf{E}_L \mid J'm')=f(J, J')(JLJ'm' \mid JL\overline{mm'-m})(L\overline{m'-m} \mid \mathbf{E}_l \mid 00)$$
$$=f^*(J', J)(J'LJm \mid J'L\overline{m'm-m'})(L\overline{m-m'} \mid \mathbf{E}_L \mid 00),$$
$$\dots \dots \dots \dots \dots \dots \dots \quad (2)$$

where $f^*(J', J)$ is unspecified but independent of m and m'. In the matrix element $(L\overline{m-m'} \mid \mathbf{E}_L \mid 00)$, the order of complexity is to $\cos^L\theta$, and therefore the model is not inconsistent in supposing that this represents the emission of a particle of spin one and orbital angular momentum L. The intensity of radiation

$$\left| (L\overline{m-m'} \mid \mathbf{E}_L \mid 00) \right|^2 = \left| (LM \mid \mathbf{E}_L \mid 00) \right|^2$$

is thus given as the sum of the squares of three terms. In the notation of equation (1) these are

$$a=(LM \mid H_L \mid 11) \quad =(L1LM \mid L1\overline{M-1}1)\alpha_{L1}Y_L^{M-1*}(\theta, \varphi),$$
$$b=(LM \mid H_L \mid 1-1)=(L1LM \mid L1\overline{M+1}-1)\alpha_{L1}Y_L^{M+1*}(\theta, \varphi),$$
$$c=(LM \mid H_L \mid 10) \quad =(L1LM \mid L1M0)\alpha_{L1}Y_L^{M*}(\theta, \varphi).$$
$$\dots \dots \dots \dots \dots \dots \dots \quad (3)$$

Using the known explicit forms of transformation coefficients given by Condon and Shortley (1935), we find

$$\left| (LM \mid \mathbf{E}_L \mid 00) \right|^2 = \mid a \mid^2 + \mid b \mid^2 + \mid c \mid^2$$
$$= \mid \alpha_{L1} \mid^2 \left\{ \frac{(L+M)(L-M+1)}{2L(L+1)} \mid Y_L^{M-1}(\theta, \varphi) \mid^2 \right.$$
$$+ \frac{(L-M)(L+M+1)}{2L(L+1)} \mid Y_L^{M+1}(\theta, \varphi) \mid^2$$
$$+ \left. \frac{M^2}{L(L+1)} \mid Y_L^M(\theta, \varphi) \mid^2 \right\}, \quad \dots \dots \dots \dots \quad (4)$$

as given by Ling and Falkoff (1949).

It might be objected that the model is not consistent with the fact that a quantum can have only two components of angular momentum about its direction of motion. This, however, is not necessarily true for the intrinsic spin of the quantum as defined here. The property of the total angular momentum L, that

$$(LM \mid \mathbf{E}_L \mid 00)_{(\theta=\varphi=0)}=0, \text{ unless } M=\pm1,$$

is not violated by equation (4).

In order to complete the method it is necessary to introduce polarization properties of the radiation field. We have found it possible to describe matrix

elements simply in terms of dipole field vectors. From a well-known property of spherical harmonics (Margenau and Murphy 1943) one may write

$$|a|^2+|b|^2+|c|^2=\left|\frac{a}{\sqrt2}e^{-i\varphi}+\frac{b}{\sqrt2}e^{i\varphi}\right|^2+\left|\left(\frac{a}{\sqrt2}e^{-i\varphi}-\frac{b}{\sqrt2}e^{i\varphi}\right)\cos\theta+c\sin\theta\right|^2.$$

$$\dots\dots\dots\dots\dots (5)$$

It is therefore consistent with equation (4) to write the matrix element for electric multipole radiation as

$$(LM\mid \mathbf{E}_L\mid 00)_{el}=\mathbf{j}\left(\frac{a}{\sqrt2}e^{-i\varphi}+\frac{b}{\sqrt2}e^{i\varphi}\right)-i\mathbf{k}\left\{\left(\frac{a}{\sqrt2}e^{-i\varphi}-\frac{b}{\sqrt2}e^{i\varphi}\right)\cos\theta+c\sin\theta\right\}$$

$$=\frac{a}{\sqrt2}(\mathbf{j}-i\mathbf{k}\cos\theta)e^{-i\varphi}+\frac{b}{\sqrt2}(\mathbf{j}+i\mathbf{k}\cos\theta)e^{i\varphi}$$

$$+c(-i\varphi_0\sin\theta). \qquad\dots\dots\dots\dots\dots\dots (6)$$

Here $\mathbf{j}$ and $\mathbf{k}$ are unit vectors in the direction in increasing θ and φ respectively. Equation (6) is in agreement with the calculations of French. In it, the co-efficients of a, b, and c are field vectors for a magnetic dipole : for magnetic multipole radiation, it would be necessary to replace them by field vectors for an electric dipole. Otherwise, equation (6), in conjunction with equations (2) and (3), gives an explicit form for a radiation matrix element of order L.

This method of separating orbital and intrinsic angular momenta is not unique. For instance, electric multipole radiation has been described (Bethe and Placzek 1937) as the emission of a particle of orbital angular momentum $(L-1)$ with intrinsic electric dipole field vector. The present system, however, has the advantage that the coefficients a, b, and c are sufficient for calculation of intensity (equation (4)) and polarization patterns (equation (6)), independently of one another.

As an example of the method, we shall calculate the matrix element $(42\mid \mathbf{E}_4\mid 00)_{el}$ for electric multipole radiation of order 4. From equation (3), with $L=4$, $M=2$, we find

$$a=\left(\frac{9}{128}\right)^{\frac12}\{3\sin\theta\,(7\cos^3\theta-3\cos\theta)\}e^{-i\varphi},$$

$$b=\left(\frac{9}{128}\right)^{\frac12}\{-7\sin^3\theta\cos\theta\}e^{-3i\varphi},$$

$$c=\left(\frac{9}{128}\right)^{\frac12}\{\sqrt2\sin^2\theta\,(7\cos^2\theta-1)\}e^{-2i\varphi}.$$

Neglecting the common numerical factor $(9/128)^{\frac12}$, the coefficients of $\mathbf{j}$ and $-i\mathbf{k}$ are

$$\frac{a}{\sqrt2}e^{-i\varphi}+\frac{b}{\sqrt2}e^{i\varphi}=\frac{4}{\sqrt2}\sin\theta\cos\theta\,(7\cos^2\theta-4)e^{-2i\varphi},$$

$$\left(\frac{a}{\sqrt2}e^{-i\varphi}-\frac{b}{\sqrt2}e^{i\varphi}\right)\cos\theta+c\sin\theta=\frac{2}{\sqrt2}\,(7\cos^2\theta-1)\sin\theta e^{-2i\varphi},$$

and equation (6) yields

$$(42\mid \mathbf{E}_4\mid 00)_{el}=\frac{1}{\sqrt2}\{\mathbf{j}\,.\,4\sin\theta\cos\theta\,(7\cos^2\theta-4)-i\mathbf{k}\,.\,2\sin\theta\,(7\cos^2\theta-1)\}e^{-2i\varphi},$$

AA

whence the intensity pattern is

$$| (42 \mid \mathbf{E}_4 \mid 00)_{el} |^2 = 2 \sin^2 \theta \ (196 \cos^6 \theta - 175 \cos^4 \theta + 50 \cos^2 \theta + 1),$$

in agreement with that calculated from equation (4).

It should be pointed out that the expression $(Jm \mid \mathbf{E}_L \mid J'm')$ introduced in equation (2) is only valid for " pure " multipole radiation of order L, defined such that its intensity pattern is the same as that of $(LM \mid \mathbf{E}_L \mid 00)$. This definition does not coincide with that given by Condon and Shortley. Thus, for instance, in the case of octupole radiation, the matrix elements of the quantity **rrr** (French 1951) lead partly to a dipole pattern, which has been suppressed (French, personal communication, 1953). This means that a dipole pattern could appear mixed with an octupole pattern even in the absence of " pure " dipole radiation, and would have the same intensity dependence on quantum energy as the accompanying octupole pattern.

III. References

BETHE, H. A., and PLACZEK, G. (1937).—*Phys. Rev.* **51** : 450.

CONDON, E. U., and SHORTLEY, G. H. (1935).—" The Theory of Atomic Spectra." (Cambridge Univ. Press.)

FALKOFF, D. L., COLLADAY, G. S., and SELLS, R. E. (1952).—*Canad. J. Phys.* **30** : 253.

FIERZ, M. (1949).—*Helv. Phys. Acta* **22** : 489.

FRENCH, A. P. (1951).—*Phil. Mag.* **42** : 263.

LING, D. S., and FALKOFF, D. L. (1949).—*Phys. Rev.* **76** : 1639.

MARGENAU, H., and MURPHY, G. M. (1943).—" The Mathematics of Physics and Chemistry." (Van Nostrand : New York.)

THE SCATTERING OF 1 MeV ELECTRONS AND POSITRONS

By J. A. McDonell*

Summary

Electrons and positrons of energy 1 MeV were selected by a β-ray spectrometer and scattered by a gold foil. The scattered particles were detected by a coincidence Geiger counter arrangement, using the same geometry for electrons and positrons. The results show that the ratio of the fraction of electrons scattered through an angle θ to the fraction of positrons scattered through θ varies with θ, and the magnitudes of the ratios agree with those predicted by theory. It is also inferred that there is a difference between the multiple scattering of electrons and positrons under the same scattering conditions.

I. INTRODUCTION

Many investigations of the single elastic scattering of electrons have been carried out since Mott (1929) calculated the cross section for the process, using the Dirac wave equation for the electron. He showed that the probability that an electron will be scattered into an element of solid angle $d\Omega$ at an angle χ to its original direction is given by

$$I(\chi)d\Omega = \varkappa^2 . R(\chi) \operatorname{cosec}^4 \tfrac{1}{2}\chi . d\Omega,$$

where

$$\varkappa^2 = \left(\frac{Ze^2w}{2mc^2(w^2-1)}\right)^2 \frac{N\sigma}{M}.$$

Z, M, and σ are the atomic number, atomic weight, and surface density in g cm^{-2} of the scatterer, and e, m, and w are the charge, mass, and total energy in mc^2 units, of the electron. N is Avogadro's number.

$\varkappa^2 \operatorname{cosec}^4 \tfrac{1}{2}\chi$ is the form of the cross section obtained from the relativistic Schrödinger equation, and is generally referred to as the " relativistic Rutherford " cross section. The important function is $R(\chi)$, which is dependent on β ($=v/c$) and α ($=Z/137$), and arises from the presence in the Dirac equation of terms which take account of the spin of the electron. For small values of α, values of $R(\chi)$ can be calculated from the first approximation

$$R(\chi) = 1 - \beta^2 \sin^2 \tfrac{1}{2}\chi + \pi\alpha\beta \sin \tfrac{1}{2}\chi . (1-\sin \tfrac{1}{2}\chi).$$

For larger values of α, $R(\chi)$ must be calculated numerically. Bartlett and Watson (1940) have performed the calculation for $Z=80$ and energies up to 1·7 MeV, and McKinley and Feshbach (1948), using an approximation in which only terms of order α^2 are retained, have calculated $R(\chi)$ for 1, 2, and 4 MeV and all values of Z. In both of these calculations, the interaction between the electron and the nucleus is represented by an unscreened Coulomb field.

* Physics Department, University of Melbourne.

For an energy of 1 MeV, with which we will be concerned, the most reliable calculation using a screened atomic field is that of Gunnersen (1952) for $Z=80$. His results show that the effect of screening is probably very small.

Massey (1942) calculated $R(\chi)$ for positron scattering, his calculations being based on those of Bartlett and Watson, with Z changed to $-Z$ throughout. He found that the values of $R(\chi)$ for electrons and positrons were quite different, and pointed out that measurements of the ratio of electron intensity to positron intensity under identical scattering conditions (hereafter called the electron-positron ratio) would provide a much more sensitive test of the theory than measurements with electrons alone. Experiments of the latter type had, up till that time, given rather conflicting results.

More recently, accurate measurements using accelerated beams of electrons have shown good agreement with theory (Van de Graaff, Buechner, and Feshbach 1946; Buechner et al. 1947), but at the time when the present work was commenced, no satisfactory measurements of the electron-positron ratio had been made, although cloud-chamber measurements by Fowler and Oppenheimer (1938) and Lasich (1948) had suggested that the ratio was of the order predicted by theory. Since then Lipkin (1952) has made accurate measurements of scattering at an angle of 58° using platinum and copper foils as scatterers and particles having various energies in the region of 1 MeV. The values of the electron-positron ratio which he obtains are generally 10 per cent. larger than the theoretical values. Howatson and Atkinson (1951) have observed the scattering of electrons and positrons by argon in a cloud chamber, and find good agreement with theory for the total number of scatters through angles greater than 20°. However, because of the small number of events involved (70 electrons and 66 positrons), their observations do not permit a significant estimate of the electron-positron ratio as a function of the scattering angle. Cusack (1952), in a similar experiment, but using positrons only, and nitrogen as the scatterer, obtained 114 scatters through angles greater than 20°. The expected number was 181, and he was unable to account for this discrepancy.

The present experiment was designed to measure the electron-positron ratio at various angles for $1\cdot07$ MeV particles scattered in gold. This energy was used by Gunnersen in his calculations.

II. Apparatus
(a) The Spectrometer

A short magnetic lens β-ray spectrometer was constructed to provide particles having energies within a small range. The instrument was similar to that of Deutsch, Elliott, and Evans (1944), and their calculations of trajectories were used in its design. From the description of the scattering arrangement which follows, it will be seen that it was desirable to focus the beam within as small an area as possible. Since the sources could not conveniently be deposited within a circle of less than $\frac{1}{4}$ in. diameter, a magnification greater than one would not have been satisfactory. On the other hand, as the magnification is

decreased, so is the transmission of the instrument, and hence the intensity of the available beam. Therefore the instrument was adjusted for unit magnification, with the central aperture fairly wide, thereby reducing the resolution in an endeavour to increase the transmission. The resolution, as measured by the width of the transmission curve at half-maximum, was 7 per cent.

During a preliminary run, it was found that particles scattered from the edges of the $\frac{1}{4}$ in. diameter exit window which ultimately defined the focused beam, were making a significant contribution to the measured count rates. This window was then machined to a diameter of $1\frac{1}{2}$ in., and other defining baffles introduced further back in the spectrometer. In this way the image of the source was confined almost entirely to a circle of $\frac{1}{4}$ in. diameter centred on the axis of the spectrometer, and in a plane $\frac{1}{8}$ in. beyond this last $1\frac{1}{2}$ in. diameter aperture. The size of the image and its location (determined by the final lining-up of the spectrometer) were examined by the use of the photographic film.

(b) The Scattering Arrangement

The arrangement of the scattering foil and the detector for the scattered particles is shown in Figure 1. The gold foil was prepared by plating gold on a thin polished copper foil, and then removing the copper with nitric acid. The foil was then mounted over one of two $\frac{7}{8}$ in. diameter holes on a vane which could be rotated by a shaft through an O-ring seal. The other hole was left blank. The foil was mounted on the side of the vane furthest from the source, and the plane of this face contained the focused image. Thus, the beam of particles could be made to pass either through the centre of the scatterer or through the centre of the blank hole. It was found that the centering of the foil or the blank hole on the spectrometer axis was best achieved by installing a small light globe inside the vacuum system, and observing the rotation of the vane visually through a piece of plate glass waxed over a small hole in the centre of the end wall of the vacuum chamber.

The scattered particles were detected by a counter whose axis was horizontal and passed through the centre of the image. This counter was mounted so that it could be rotated about a vertical axis, also through the centre of the image, the rotation being effected by an O-ring seal. A pointer rotating with the counter but outside the vacuum system moved across an angular scale, and provided a measure of θ, the angle which the counter axis made with the spectrometer axis, to within $\pm\frac{1}{4}°$.

In the design of the counter, the main consideration was reduction of the high background produced by the energetic γ-rays from the positron source. It was found that these were strongly scattered in the magnet coil of the spectrometer, producing a high γ-ray flux in the region of the detector. This was no doubt increased by a certain amount of annihilation radiation from positrons striking the baffles in the spectrometer.

Accordingly, a counter had to be designed having a high directional sensitivity. The form finally chosen is shown in Figure 2. The counter is of the Maze type, with a body of high conductivity soda glass, and an external

"Aquadag" cathode. The cathode is "split" into three sections, and the centre section is at the anode potential, so that a discharge at one end of the

HORIZONTAL SECTION   A

1 0 1 2 3 4 5 6 IN

VERTICAL SECTION A-A
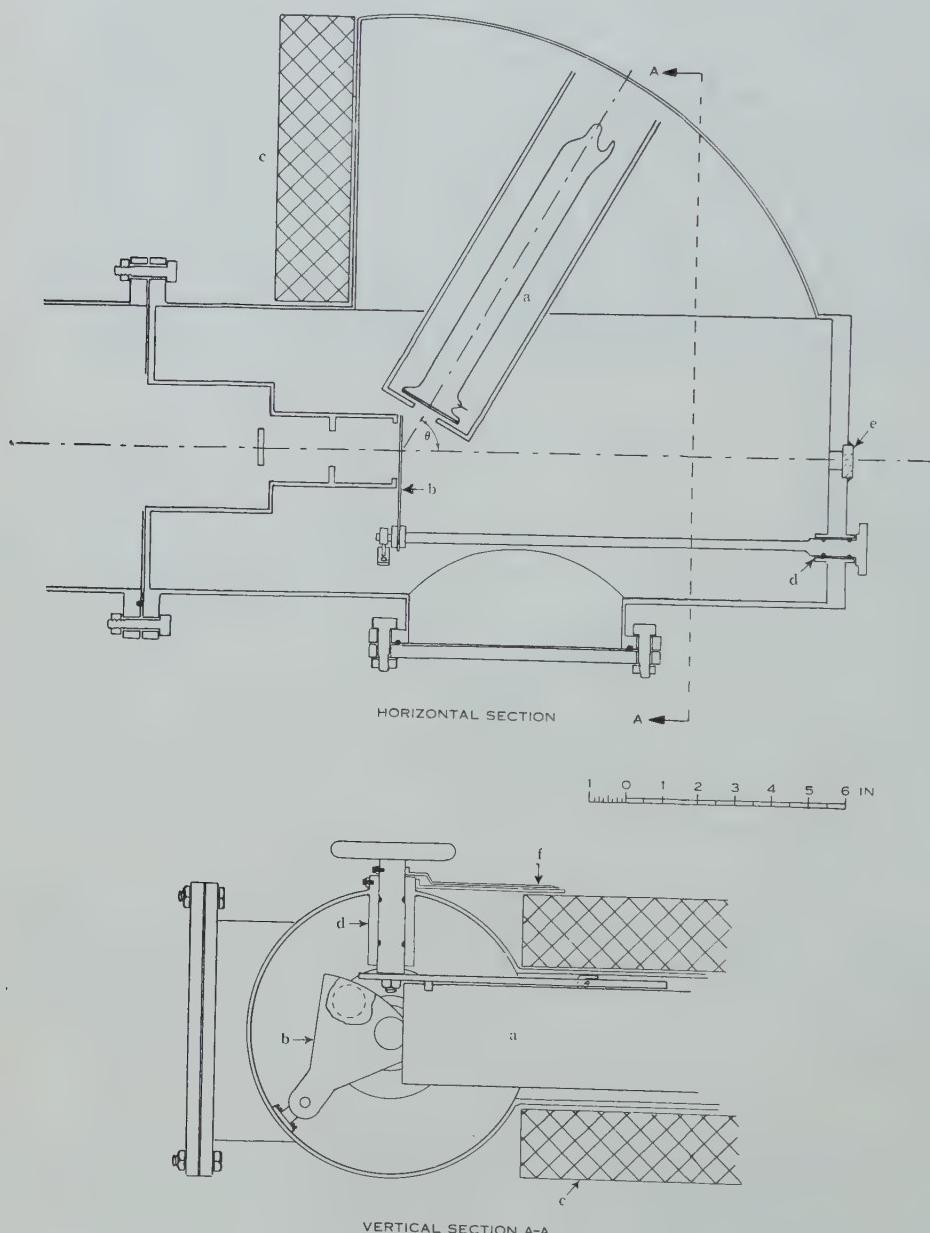
Fig. 1.—Scattering arrangement. *a*, Counter in brass box ; *b*, vane carrying gold foil ; *c*, lead shielding ; *d*, O-ring seals ; *e*, glass window ; *f*, pointer and angular scale.

counter cannot travel to the other end.   Positive pulses are taken off each of the two outside cathode sections, and coincidences between them counted.   Thus a

single particle is counted only when it traverses practically the whole counter, and its path makes only a fairly small angle with the counter axis. As an indication of the effectiveness of this arrangement, its cosmic ray background with the axis horizontal was $0 \cdot 5$/min. The conductivity of the soda glass was high enough to prevent any appreciable decrease in pulse height as the count rate increased. To reduce further the γ-ray background, the quadrant in which the counter rotated (Fig. 1) was lead-shielded above, below, and on the side nearest the magnet coil.

The end window of the counter was made of soda glass, sealed on to the body with solder glass. The window was thick enough to absorb electrons of energy less than $0 \cdot 85$ MeV, this energy limit being checked using the direct beam of focused electrons in the spectrometer. Thus any particles scattered inelastically through the angles at which measurements were taken would not be detected, nor would any " stray " low-energy particles such as were observed by Schulze-Pillot and Bothe (1950) in their spectrometer.
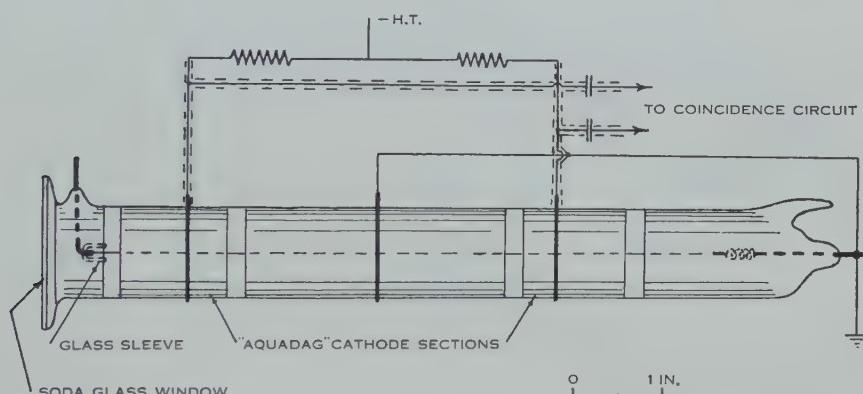
Fig. 2.—End window directional counter.

There is reason to expect that this absorber would affect identical beams of electrons and positrons rather differently. However, the relative strengths of the electron and positron beams incident on the scatterer were measured using this same counter, in the manner described below. Therefore, since the quantities which we measure have the form

$$\frac{\text{Fraction of incident electrons scattered through } \theta}{\text{Fraction of incident positrons scattered through } \theta},$$

the presence of an absorber will not affect the results unless the energy distributions of the *scattered* electrons and positrons differ as a result of their having traversed the gold foil. There is no reason to believe that positrons differ greatly from electrons in the way they lose small amounts of energy, and since the theory of Landau (1944) gives the most probable energy loss for electrons in this case as $0 \cdot 6$ per cent. of the initial energy, one can assume that the effect of any such difference in our measurements is negligible.

(c) Sources*

The energy calibration of the spectrometer was obtained using the photo-electrons produced in Pb by the γ-rays of ⁶⁰Co. The energies of these γ-rays have been accurately measured by Lind, Brown, and Du Mond (1949).

The sources of positrons and electrons had to be chosen subject to the fairly severe limitations of having a spectrum with maximum energy rather greater than 1 MeV, and a half-life of at least several days, since all radioactive materials had to be imported. Of the possible electron sources, the most readily available was ³²P (obtained from Harwell) and a 20 mc source of this isotope was found to be quite satisfactory. A further advantage is that ³²P has no γ-rays, and therefore the background count is at a minimum. For a positron source, however, there is no choice, as the only available positron emitter with suitable energy and half-life is ⁵⁶Co. While this isotope has a maximum positron energy of 1·5 MeV and a half-life of 75 days, it has certain less desirable features from our point of view.

TABLE 1

ISOTOPES PRESENT IN A SOURCE OF RADIO-COBALT

Isotope	⁵⁶Co	⁵⁷Co	⁵⁸Co
Half-life (days)	75	270	75
Decay particles	β^+, 0·995	β^+, 0·315	β^+, 0·45
(max. energy MeV) ..	β^+, 1·53		
γ-Ray energies (MeV) ..	0·845	0·119	0·805
	1·26	0·131	
	1·74		
	2·01		
	2·55		
	3·25		
Production	⁵⁶Fe-d-2n	⁵⁶Fe-d-n	⁵⁷Fe-d-n

⁵⁶Co is produced by deuteron bombardment of Fe or Ni, or by α-particle bombardment of Mn, but in all cases, other cobalt isotopes are formed as well. The most recent investigation of these isotopes is that of Cheng, Dick, and Kurbatov (1952). In our case, the radio-cobalt was produced by deuteron bombardment of Fe†, and the main isotopes present are those shown in Table 1. After the conclusion of the scattering measurements, the β- and γ-spectra of the cobalt source were investigated. The spectra obtained were in good agreement with those previously reported.

Although no published data are available, one would assume that the Fe-d-n reaction would be more probable than the Fe-d-2n reaction, so that a

* All radioactive materials used in this work were obtained through the agency of the Tracer Elements Section, C.S.I.R.O.

† We are very grateful to the Department of Terrestrial Magnetism, Washington, U.S.A., which prepared this source.

cobalt source of given strength would contain more ^{57}Co than ^{56}Co. Further, if the strength of the source is estimated from γ-ray intensity measurements, it is likely that the measured intensity will be largely due to the 119 keV and 131 keV γ-rays of ^{57}Co. It is assumed that effects such as this account for the fact that the number of 1 MeV electrons from 20 mc of ^{32}P was 160 times greater than the number of 1 MeV positrons from our 12 mc source of radio-cobalt. (These source strengths are those quoted by Harwell and the Department of Terrestrial Magnetism.)

III. EXPERIMENTAL PROCEDURE

The surface density of the foil was measured as follows. A piece of foil (about 5 cm^2) from the same sample from which the scatterer was taken was placed on a piece of glass and illuminated from below. It was then covered by a piece of graph paper having millimetre divisions of good precision. The outline of the piece of foil was carefully traced on the graph paper and its area estimated by counting squares. The piece was then weighed. It was then divided in two and the procedure repeated for each small piece. The surface densities thus obtained were : for the whole piece, 9·65 mg cm^{-2} and, for the two small pieces, 9·65 and 9·67 mg cm^{-2}. On the basis of these measurements, the foil was taken to be sufficiently uniform, and of surface density 9·65 mg cm^{-2}.

With the counter set at a particular angle, and the beam passing through the blank hole, the count rate was made up of contributions from :

(a) cosmic radiation,

(b) γ-ray background (for the positron source only),

(c) electrons scattered into the counter from baffles etc. in the apparatus.

With the foil brought into position, the count rate was made up of (a) and (b), together with electrons scattered by the foil, and some fraction of (c). This fraction was unknown, unless one could in some way determine whether the source of these particles lay in front of, or behind the scatterer. Thus the baffles were located as described above in order to reduce (c) to negligible proportions. Under these conditions, the count rate due to electrons scattered by the foil is just the difference between the rates with the foil " in " and " out ".

As a check on the effectiveness of the baffle arrangement, the count rate as a function of θ with the foil out was measured before the scattering measurements were commenced, using a ^{32}P source. Since this source emits no γ-rays, all electrons counted in this measurement must have been in category (c). The result is shown in Figure 3, and indicates that satisfactory measurements of the particles scattered from the foil could be expected for $\theta \geqslant 40°$. The " plateau " at about 10° on the curve in Figure 3 is due to particles entering the counter from the direct beam of the spectrometer. The intensities of the positron and electron beams incident on the foil were compared by measuring the count rates at $\theta = 10°$, with the foil out.

While measurements of the electron-positron ratio would not be affected by any variation with θ of the effective solid angle subtended at the centre of the scatterer by the counter, measurements of the relative intensities at different

angles for electrons or positrons alone are of some interest.   For such measure-
ments, some information about this solid angle is necessary.   A thin (less than
0·1 mg cm⁻²) " Formvar " film was temporarily mounted over the blank hole
in the vane, and a small amount of ^{32}P evaporated at its centre.   This was then
lined up on the spectrometer axis, providing a source of electrons isotropically
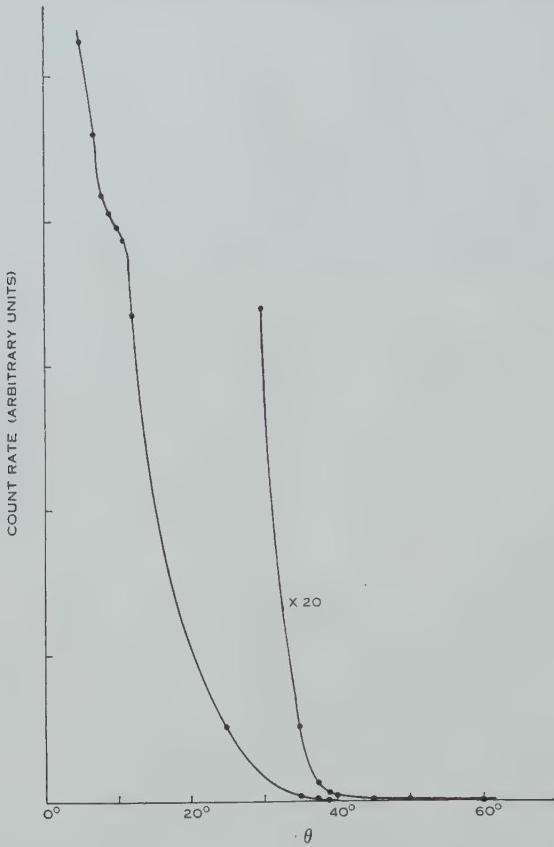
Fig. 3.—Count rate v. angle with foil out and ^{32}P source.

distributed in space, except for directions making small angles with the plane of
the film, where some effect due to backscattering might have been expected.
The count rate under these conditions was constant for $\theta \leqslant 70°$.   Thus the
angular range in which scattering measurements were made was $40° \leqslant \theta \leqslant 70°$.

For the scattering measurements using positrons, the following procedure
was adopted.   At the beginning of each day, a count was taken in the 10°,
foil out position.   This served to monitor the decay of the source, and also to
check that the apparatus was operating correctly.   Four-hour counts were then
taken in the foil out and foil in positions for a certain angle, say 70°.   On subse-
quent days, these measurements would be made at 60°, 50°, 40°, 70°, 60°, . . . ,
and so on.   Counting at the lower angles was discontinued when sufficient
accuracy had been attained.

With the electron source, the count rate in the 10°, foil out position was initially so high that an appreciable fraction of counts was lost due to the dead time of the counter. The count rate in the 40°, foil in position was used instead to monitor the decay of the source. One-hour counts in the foil out and foil in positions were taken at each angle, and the angles taken in turn as previously. The count rate in the 10°, foil out position was taken daily, and its initial value calculated in the following way.

Let $R(t)$ be the count rate at time t, then

$$R(t) = R(0)e^{-\lambda t}(1 - R(0)e^{-\lambda t}\tau),$$

where τ is the dead time of the counter and λ is known from other measurements on the decay of the source.

Then

$$\ln R(t) = \ln R(0) - \lambda t + \ln (1 - R(0)e^{-\lambda t}\tau)$$
$$\simeq \ln R(0) - \lambda t - R(0)e^{-\lambda t}\tau \text{ (since } \tau \text{ is small)},$$

and

$$\frac{\mathrm{d}}{\mathrm{d}t} \ln R(t) = -\lambda(1 - R(0)e^{-\lambda t}\tau).$$

Thus, by plotting $\ln R(t)$ against t and measuring the slope of the curve, the correction term $R(0)e^{-\lambda t}\tau$ could be found, and hence $R(0)$. The correction term was quite small, being $0 \cdot 038$ at $t = 20$ days.

The cosmic ray background was measured and found to be independent of angle, within the statistical errors of the measurements.

The scattered intensities at the various angles had to be determined from measurements taken over about one half-life of each source. To do this, the decay was neglected over the period of each pair of foil out and foil in counts, so that for a given angle, if t_i is the time at the change over from foil out to foil in, we have

x_i = the count recorded during the standard time interval at t_i with the foil out,

y_i = the corresponding count with the foil in.

Then, if μ_c = the mean rate due to cosmic rays,

μ_γ = the mean initial rate due to γ-rays from the source, and

μ_e = the mean initial rate due to electrons scattered by the foil,

x_i is Poisson distributed with mean

$$\mu_c + \mu_\gamma e^{-\lambda t_i},$$

and y_i is Poisson distributed with mean

$$\mu_c + (\mu_\gamma + \mu_e)e^{-\lambda t_i}.$$

λ and μ_c are known, and we want to estimate μ_e and the variance in our estimate of μ_e. Using a least squares method, we form the sum

$$S = \sum_i [\{x_i - \mu_c - \mu_\gamma e^{-\lambda t_i}\}^2 + \{y_i - \mu_c - (\mu_\gamma + \mu_e)e^{-\lambda t_i}\}^2],$$

and from the equations

$$\frac{\partial S}{\partial \mu_e} = 0, \qquad \frac{\partial S}{\partial \mu_\gamma} = 0,$$

we obtain

$$\hat{\mu}_e = \frac{\Sigma(y_i - x_i)e^{-\lambda t_i}}{\Sigma e^{-2\lambda t_i}}$$

as our estimate of μ_e,

$$\mu\gamma = \frac{\Sigma x_i e^{-\lambda t_i} - \mu_c \Sigma e^{-\lambda t_i}}{\Sigma e^{-2\lambda t_i}},$$

and

$$\text{var } (\hat{\mu}_e) = \frac{\Sigma[E(x_i) + E(y_i)]e^{-2\lambda t_i}}{(\Sigma e^{-2\lambda t_i})^2},$$

where

$$E(x_i) = \mu_c + \hat{\mu}\gamma e^{-t_i},$$

and

$$E(y_i) = \mu_c + (\hat{\mu}\gamma + \hat{\mu}_e)e^{-t_i}.$$

The minimum variance in the result should be obtained if one uses a maximum likelihood rather than a least squares technique. This was done for one case, but the difference was slight, and the simpler least squares method was used throughout.

For the electron measurements, the standard errors, defined by

$$\sigma(\hat{\mu}_e) = [\text{var } (\hat{\mu}_e)]^{\frac{1}{2}}/\hat{\mu}_e,$$

were all less than 1 per cent., and x_i and y_i were used instead of $E(x_i)$ and $E(y_i)$ in the calculation of var $(\hat{\mu}_e)$, in which case values of $\hat{\mu}\gamma$ were not required.

The results of these calculations are shown in Table 2.

IV. COMPARISON WITH THEORY

In order to obtain usable count rates with the positron source available, the foil used was rather thicker than originally intended. As a result, the intensities of the scattered particles contained large contributions due to multiple scattering.

The usual procedure in single scattering experiments is to make corrections for multiple scattering using a technique such as that of Chase and Cox (1940) or of Butler (1950). In our case, however, the corrections obtained in this way were so large that they could not be relied upon. What was required was a complete angular distribution, covering in particular the region of transition from multiple scattering at small angles to essentially single scattering at large angles. Further, separate distributions of this kind for electrons and positrons were needed. The only existing multiple scattering theory which produces such a complete distribution in a readily calculable form is that of Molière (1948), but his theory makes no distinction between electrons and positrons. However, it has been shown by the experiments of Groetzinger, Humphrey, and Ribe (1952) and Seliger (1952) that there is a difference in the multiple scattering of these particles.

Thus a technique was devised for producing the separate electron and positron distributions. Firstly, Molière's complete distribution $f(\chi)$* was calculated up to $\chi = 120°$, taking the effective thickness of the foil to be its actual

* Molière (1948), equation (7·3), p. 88.

thickness divided by cos 11°. This angle of 11° was the mean angle that the incident particles made with the normal to the foil, as estimated from the calculated trajectories in the spectrometer. Now, for large values of χ, $f(\chi)$

TABLE 2

EXPERIMENTAL RESULTS

Positrons			
θ	$\hat{\mu}_\gamma$ (counts/4 hr)	$\hat{\mu}_e$ (counts/4 hr)	$\sigma(\hat{\mu}_e)$ (%)
40°	296·6	1448	2·45
50°	265·2	416	5·22
60°	167·1	168·5	7·60
70°	146·8	45·5	14·60

Electrons		
θ	$\hat{\mu}_e$ (counts/hr)	$\sigma(\hat{\mu}_e)$ (%)
40°	62800	0·40
45°	39600	0·55
50°	25800	0·63
55°	16760	0·68
60°	11590	0·68
65°	7570	0·81
70°	5150	0·72

10°, foil out rates : 100·6/min (positrons), 15780/min (electrons). From these figures, we calculate $R_E(\theta)$ = ratio of electron intensity to positron intensity.

θ	$R_E(\theta)$	Standard Error (%)
40°	1·11	2·5
50°	1·58	5·3
60°	1·75	7·6
70°	2·88	14·6

asymptotes as expected to the single scattering distribution from which it is generated. However, instead of the exact single scattering distribution

$$I(\chi) = \varkappa^2 \operatorname{cosec}^4 \tfrac{1}{2}\chi \cdot R(\chi), \qquad \cdots\cdots\cdots\cdots \quad (1)$$

Molière uses the approximate form

$$I'(\chi) = 16\chi^2/\chi^4.$$

For large values of χ, this is a poor approximation, and the result is that, for electrons, $f(\chi)$ lies well below the exact single scattering distribution at large angles, as is shown in Figure 4.
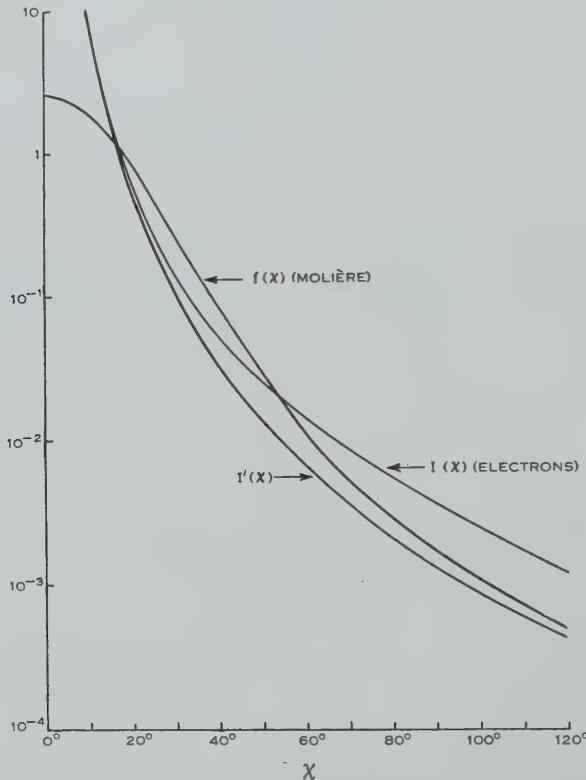
Fig. 4.—Relation between multiple and single scattering distributions.

A new distribution was then formed by calculating

$$f'(\chi) = I(\chi)[f(\chi)/I'(\chi)].$$

The implication here is that the ratio $f(\chi)/I'(\chi)$ is regarded as giving the relation between the single scattering distribution and the multiple scattering distribution resulting from it.

Next, the function

$$y(\chi) = [f'(\chi) - f(\chi)] \cdot 2\pi \sin \chi$$

was calculated.

Now our function $f'(\chi)$ is un-normalized. Since $f'(\chi)$ is very small indeed for angles greater than 120°, we can normalize it by multiplying by $1/k'$, where

$$k' = \int_0^{2\pi/3} f'(\chi) \cdot 2\pi \sin \chi d\chi.$$

However, we know what values our final distribution should take for angles near 120°, because at these angles we find that the Chase and Cox correction for multiple scattering* is less than 15 per cent. and therefore can be taken as giving fairly accurately the corrections to be applied to $I(\chi)$. Denoting the Chase and Cox correction by $[1+C(\chi)]$, we now have to modify slightly our function $f'(\chi)$, so that it becomes $f_n(\chi)$, with

$$\frac{1}{k}f_n\,(120°)=I\,(120°)[1+C\,(120°)],$$

where

$$k=\int_0^{2\pi/3} f_n(\chi)\,.\,2\pi\,\sin\,\chi\mathrm{d}\chi.$$

We can write

$$f'(\chi)=f(\chi)+y(\chi)/2\pi\,\sin\,\chi,$$

and so the modification can be made by replacing $y(\chi)$ by a new function $y_n(\chi)$, such that

$$f_n(\chi)=f(\chi)+y_n(\chi)/2\pi\,\sin\,\chi,$$

where

$$\frac{f\,(120°)+y_n\,(120°)/2\pi\,\sin\,120°}{\int_0^{2\pi/3} f(\chi)\,.\,2\pi\,\sin\,\chi\mathrm{d}\chi+\int_0^{2\pi/3} y_n(\chi)\mathrm{d}\chi}=I\,(120⁰)[1+C\,(120°)].\;\ldots\;(2)$$

A slightly different technique for choosing $y_n(\chi)$ was adopted for the cases of electrons and of positrons. In the electron case, $y(\chi)$ was found to be well fitted throughout the range by a function of the form

$$y(\chi)=a(1-\mathrm{e}^{-b\chi^2})\mathrm{e}^{-c\chi}.$$

Then the parameters a, b, and c were varied by trial until condition (2) was satisfied. For the positron case, it was found that plotting $\log y(\chi)$ against χ gave practically a straight line from $\chi=60°$ onwards. Suitable values of $y_n(\chi)$ were obtained by replacing this straight section by the tangent to the curve at an angle slightly less than 60°.

The whole calculation was designed to produce distribution curves which only differed slightly from the general form of the Molière distribution, and the final normalized distributions are shown in Figures 5 (a) and 5 (b).

The justification for the somewhat arbitrary procedure adopted is that the resultant distributions have the following features :

(1) In the multiple scattering region, they have very much the same form as that of Molière, and in fact differ from the latter by 10 per cent. or less for angles up to 35°. This similarity in form is regarded as desirable, in view of the recent measurements of Hanson et al. (1951), who obtain good agreement with

---

* The second order correction

$$\frac{\overline{\Delta I(\chi)}}{I(\chi)}=\overline{\varepsilon^2}\,(\mathrm{cosec}^2\,\tfrac{1}{2}\chi-\tfrac{1}{2})+\overline{\varepsilon^4}\left(\frac{9}{16}\,\mathrm{cosec}^4\,\tfrac{1}{2}\chi-\frac{11}{24}\,\mathrm{cosec}^2\,\tfrac{1}{2}\chi+\frac{1}{24}\right)$$

has been used here. This expression was derived by Lipkin (1950), but with some numerical errors in the coefficients.

Molière's distribution for a case in which the mean angle of the distribution is of the order of 3°, i.e. the approximate single scattering cross section can be regarded as valid.
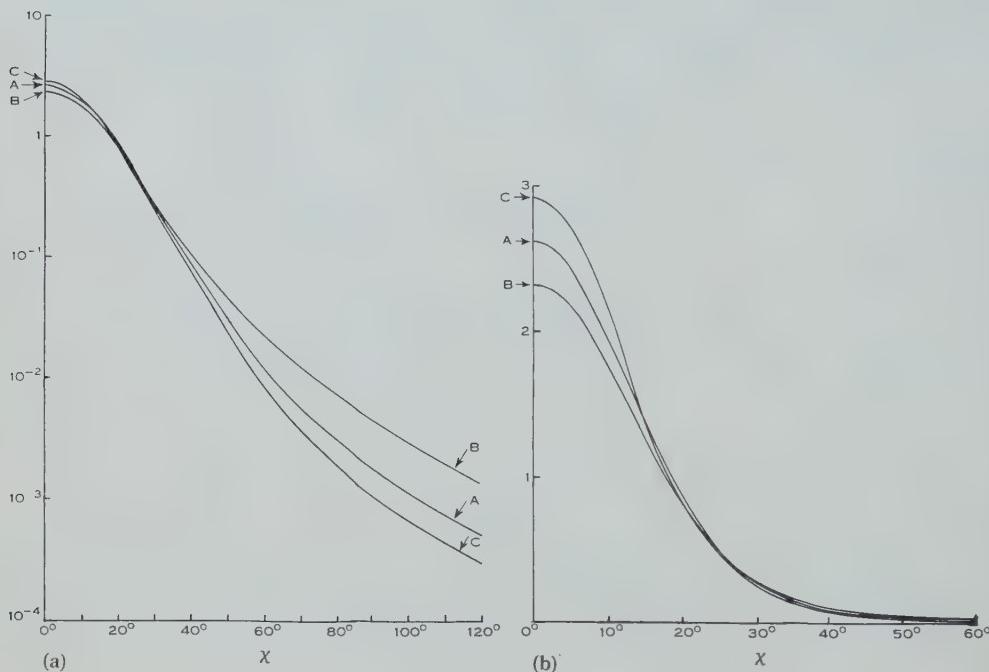
Fig. 5.—Normalized complete angular distributions shown on logarithmic (a) and linear scales (b). Curve A, original Molière distribution, $f(\chi)$; curve B, $(1/k)f_n(\chi)$ for electrons; curve C, $(1/k)f_n(\chi)$ for positrons.

(2) At large angles, the distributions take the values predicted by the accurate single scattering theory, corrected for the effect of multiple scattering. It is shown in Table 3 that, while the distributions were obtained by forcing a fit at 120°, the agreement is good over a considerable angular range.

TABLE 3

COMPARISON OF VALUES OF THE CORRECTED SINGLE SCATTERING DISTRIBUTIONS WITH THE VALUES OF OUR FINAL DISTRIBUTIONS

χ	Electrons		Positrons	
	$\dfrac{1}{k}f_n(\chi)$	$I(\chi)[1+C(\chi)]$	$\dfrac{1}{k}f_n(\chi)$	$I(\chi)[1+C(\chi)]$
90°	0·00466	0·00465	0·001090	0·001083
100°	0·00299	0·00296	0·000670	0·000663
110°	0·00199	0·00196	0·000440	0·000440
120°	0·00134	0·00134	0·000300	0·000300

(3) The mean angles of the distributions are $22 \cdot 9°$ for electrons and $19 \cdot 0°$ for positrons, a difference which is consistent with the experimental results which compare the multiple scattering of electrons and positrons. Groetzinger, Humphrey, and Ribe (1952) report an r.m.s. angle for positrons 10 per cent. less than that for electrons scattered in nitrogen. Lasich (1948) obtained a smaller mean angle for positrons than for electrons with gold as the scatterer, although his results are hardly significant in this respect.
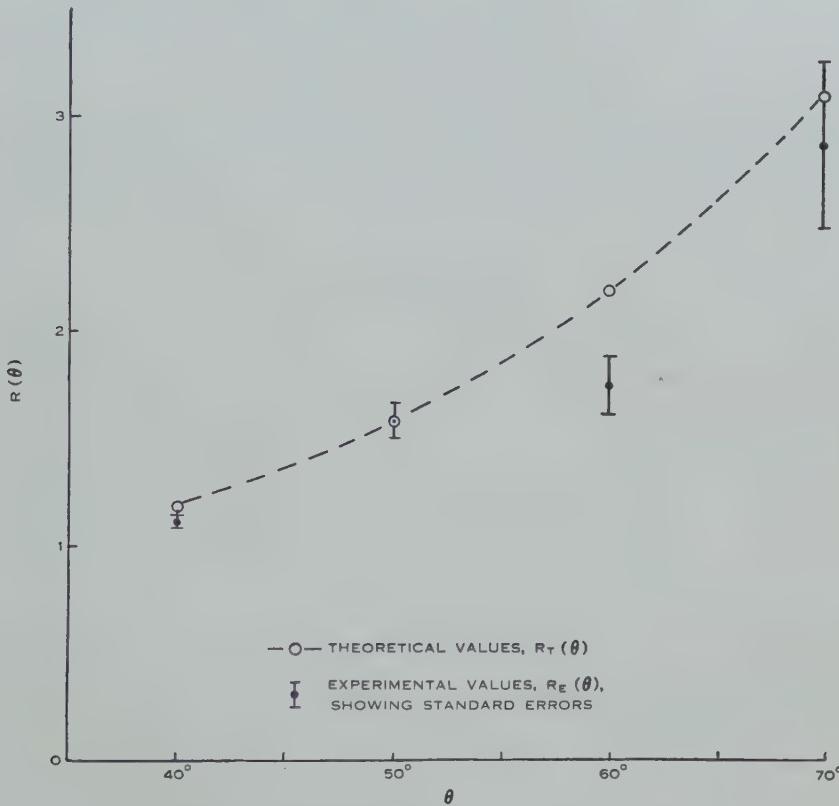
Fig. 6.—Values of the electron-positron ratio.

Finally, to calculate the expected scattered intensities in our apparatus, the incident beam was regarded as being uniformly distributed round the surface of a cone of half-angle α, in which case, the fraction of the beam scattered into an element of solid angle $d\Omega$ at an angle θ in the apparatus is

$$P(\theta)d\Omega = \frac{d\Omega}{2\pi} \int_0^{2\pi} \frac{1}{k} f_n(\chi) d\varphi,$$

where $\cos \chi = \cos \alpha \cos \theta + \sin \alpha \sin \theta \cos \varphi$.

All the integrations in the above work were, of course, carried out numerically.

For comparison with the experimental results, we calculate

$$R_T(\theta) = P(\theta)^- / P(\theta)^+,$$

the superscripts referring to electrons and positrons. These are compared
with the values $R_E(\theta)$ from Table 2, in Figure 6.

Also the ratios $\rho(\theta)=\hat{\mu}_e(\theta)/P(\theta)$ were calculated for electrons and positrons,
and these are shown in Figure 6.
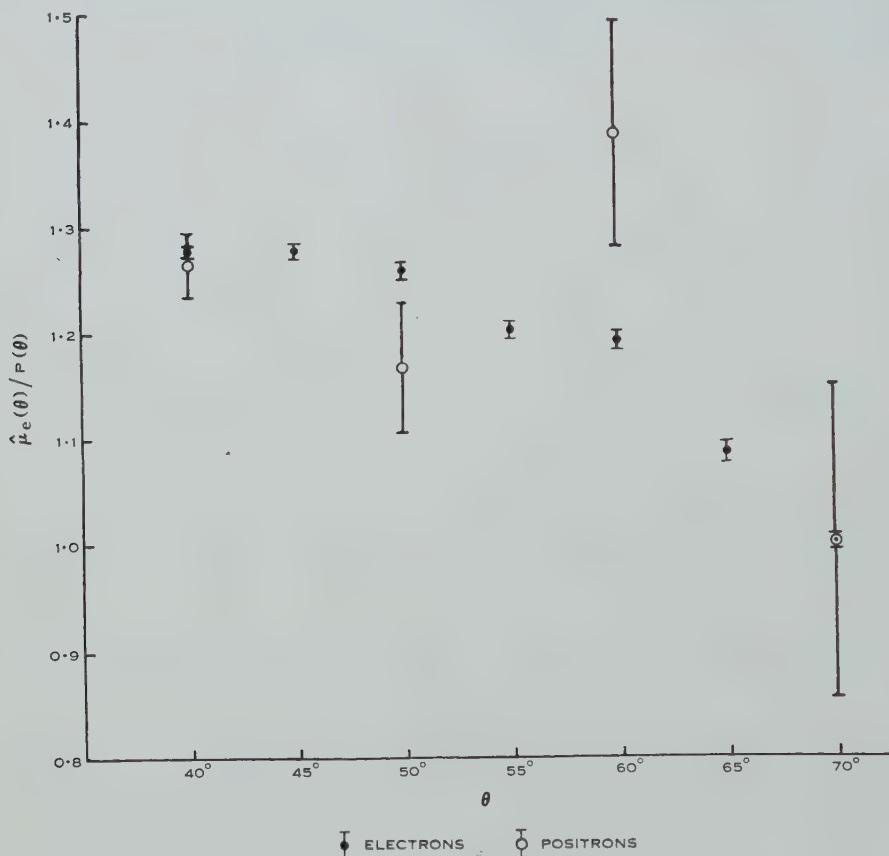
Fig. 7.—Ratios of experimental and theoretical intensities. The ratios have been
multiplied by the appropriate factors to reduce the 70° value to unity in each case.

V. Discussion

The experimental results clearly establish a variation of the electron-positron
ratio with the angle of scattering. Whether or not the measured values of the
ratio can be said to agree with the theoretical values, depends on how much
reliance can be placed on the calculations outlined in Section IV. From Figure 7,
one deduces that, over this angular range, our method of calculation may have
produced incorrect predicted intensities. However, since the errors introduced
are, within the limits of experimental error, the same fraction of the intensity
for both electrons and positrons at any angle, the electron-positron ratios as
calculated will be unaltered by such errors.

If we take this to be the case, and it seems reasonable to do so, then our experimental values of the electron-positron ratio agree well with theory, and confirm Gunnersen's (1952) calculation which predicts no significant effect on the cross section due to the screening of the nucleus.

Also, since the mean angles of the complete angular distributions are approximately 20°, the existence of an electron-positron ratio for angles of the order of twice the mean leads to the conclusion that electrons and positrons must have different multiple scattering distributions. This is not predicted by any of the existing theories of multiple scattering.

VI. Acknowledgments

VII. References

BARTLETT, J. H., JR., and WATSON, R. E. (1940).—*Proc. Amer. Acad. Arts Sci.* **74** : 53.

BUECHNER, W. W., VAN DE GRAAF, R. J., SPERDUTO, A., BURRILL, E. A., and FESHBACH, H. (1947).—*Phys. Rev.* **72** : 678.

BUTLER, S. T. (1950).—*Proc. Phys. Soc. Lond.* A **63** : 599.

CHASE, C. T., and COX, R. T. (1940).—*Phys. Rev.* **58** : 243.

CHENG, L. S., DICK, J. L., and KURBATOV, J. D. (1952).—*Phys. Rev.* **88** : 887.

CUSACK, N. (1952).—*Phil. Mag.* **43** : 671.

DEUTSCH, M., ELLIOTT, L. G., and EVANS, R. D. (1944).—*Rev. Sci. Instrum.* **15** : 178.

FOWLER, W. A., and OPPENHEIMER, J. (1938).—*Phys. Rev.* **54** : 320.

GROETZINGER, G., HUMPHREY, W., JR., and RIBE, F. L. (1952).—*Phys. Rev.* **85** : 78.

GUNNERSEN, E. M. (1952).—*Aust. J. Sci. Res.* A **5** : 258.

HANSON, A. O., LANZL, L. H., LYMAN, E. M., and SCOTT, M. B. (1951).—*Phys. Rev.* **84** : 634.

HOWATSON, A. F., and ATKINSON, J. R. (1951).—*Phil. Mag.* **42** : 1186.

LANDAU, L. (1944).—*J. Phys. Moscow* **8** : 201.

LASICH, W. B. (1948).—*Aust. J. Sci. Res.* A **1** : 249.

LIND, D. A., BROWN, J. R., and DU MOND, J. W. (1949).—*Phys. Rev.* **76** : 1838.

LIPKIN, H. J. (1950).—Ph.D. Thesis, Princeton University.

LIPKIN, H. J. (1952).—*Phys. Rev.* **85** : 517.

McKINLEY, W. A., JR., and FESHBACH, H. (1948).—*Phys. Rev.* **74** : 1759.

MASSEY, H. S. W. (1942).—*Proc. Roy. Soc.* A **181** : 14.

MOLIÈRE, G. (1948).—*Z. Naturf.* **3a** : 78.

MOTT, N. F. (1929).—*Proc. Roy. Soc.* A **124** : 425.

SCHULZE-PILLOT, G., and BOTHE, W. (1950).—*Z. Naturf.* **5a** : 440.

SELIGER, H. H. (1952).—*Phys. Rev.* **88** : 408.

VAN DE GRAAFF, R. J., BUECHNER, W. W., and FESHBACH, H. (1946).—*Phys. Rev.* **69** : 452.

B

THE DISTRIBUTION OF RADIO BRIGHTNESS OVER THE SOLAR DISK AT A WAVELENGTH OF 21 CENTIMETRES

II. THE QUIET SUN—ONE-DIMENSIONAL OBSERVATIONS

By W. N. Christiansen* and J. A. Warburton*

Summary

Daily records of one-dimensional distribution of radio brightness over the Sun are obtained in the way described in Part I of this series (Christiansen and Warburton 1953). When superimposed, these records show a well-marked lower envelope which remains substantially the same, in shape and size, over a period of months. This envelope gives the brightness distribution over the " quiet " Sun. The direction of scan, with respect to the position of the solar axis, changes during a year by more than 50°. The very small change in shape of the envelope during this period suggests that for purposes of preliminary calculation the brightness distribution may be assumed to be circularly symmetrical. The radial distribution of brightness, calculated on this basis, is found to show marked limb-brightening and to be consistent with the calculated brightness distribution for a simple solar model in which the assumed values of temperature and density are close to those commonly accepted.

I. INTRODUCTION

The Sun is known to show limb-darkening when viewed optically. This is attributed to the negative temperature gradient outwards in the photosphere. What might be expected at radio frequencies cannot be determined from temperature gradients alone, because one cannot assume that all rays reach effectively infinite optical depth (Smerd 1950). However, the high temperature of the corona relative to the inner atmosphere of the Sun, derived from optical evidence and confirmed by radio observation (Pawsey 1946 ; Pawsey and Yabsley 1949), is likely to lead to limb-brightening at some radio frequencies. This was first suggested by Martyn (1946, 1948). More detailed calculations by Unsöld (1947), Waldmeier and Müller (1948), Smerd (1950), Denisse (1950), and Reule (1952) have shown the magnitude of limb-brightening to be expected from various " models " of the Sun. All have indicated marked limb-brightening at some part of the centimetre-decimetre range of wavelengths.

Much information on electron temperatures and densities in the chromosphere and corona, and particularly in the relatively unknown transition region between these, could be gained from observed brightness distributions at different radio frequencies. Many attempts have been made to derive the brightness distribution, usually during solar eclipses. Results to date have been disappointing, mainly because of the difficulty of separating the effects of small

---

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

areas of enhanced radio brightness (usually associated with sunspots) from the background or " quiet-Sun " radiation.   This difficulty arises because in a single set of observations, such as that obtained during a solar eclipse, variable and constant components cannot be distinguished from each other.   In addition, the reduction of eclipse observations to a one-dimensional or a radial brightness distribution presents considerable, if not insurmountable, difficulties.   It is clearly desirable to obtain a brightness distribution of simple geometry and high resolution and capable of being separated into steady background and short-lived components.

In an attempt to measure the brightness distribution over the solar disk in the absence of variable components, Stanier (1950) recorded the radiation from the Sun during periods when the Sun was relatively free from areas of enhanced radio brightness.   A two-aerial interferometer was used and the amplitude of the interference pattern was determined for different spacings between the aerials.   From these a one-dimensional brightness distribution over the solar disk was derived and, by an assumption of circular symmetry, the radial brightness distribution was determined.   Stanier found no sign of the expected limb-brightening at a wavelength of 60 cm.

The results obtained, however, must be treated with some reserve since the relative phases of the Fourier components were not determined and it was necessary therefore to assume that the one-dimensional brightness distribution was symmetrical.   The published experimental curves show the presence of localized bright areas, which are probably situated asymmetrically with respect to the centre, and these could cause an error in the final result.

A more direct way of determining the one-dimensional brightness distribution is obviously to be preferred.   The 32-element interferometer described in Part I of this series (Christiansen and Warburton 1953), which has a beam width in an east-west direction of only 3′ of arc, has been employed for this purpose at a wavelength of 21 cm.

II. Observations

As described in Part I, daily records of the one-dimensional brightness distribution across the solar disk are obtained.   These daily records normally show peaks of intensity which change in position from day to day and indicate the effects of areas of enhanced radio brightness on the disk.   The change, with time, in the shape of the brightness records makes possible an investigation of the existence of a steady base-level component of the solar radiation and a determination of its brightness distribution.

The method adopted was to superimpose a number of daily records and investigate the lower boundary of these curves.   Before this could be done, however, it was necessary to ensure that the curves were all on the same scale and to find reference axes by means of which the curves could be superimposed. Changes in the gain of the system from one day to the next were detected, as stated in Part I, by a daily measurement of the area of the interferometer record produced by the Sun's passage through the aerial beam and a comparison of this with the reading of an instrument which registered the magnitude of the

BB

radio energy received from the whole solar disk. The result could be checked by comparison of suitable parts of the interferometer records on successive days. Hence short-term changes in the vertical scale of the records could be detected and corrected.

Short-term changes in the horizontal (time) scale of the records were small and could be detected by measuring the distance between successive time marks on the records. Longer-term changes will be mentioned later.

In superimposing records for different days, the horizontal and vertical registration lines are respectively the line drawn by the receiver when the aerial beams are directed away from the Sun and a line drawn at the instant when the centre of the solar disk passes through an aerial beam. The position of this second line on the record can be found from calculations of the directions of the aerial beams and from astronomical data relating to the Sun.

Calculations of the relative positions of the Sun and aerial beams were carried out for several weeks and lines were drawn on the interferometer records corresponding to the times of passage of the centre of the solar disk through the aerial beams. It soon became apparent that these lines always fell very close to the centre of the patterns on the interferometer records, the maximum difference corresponding to a time interval of less than 4 sec. This indicated that the central lines of the optical and radio disks could be considered identical for the purposes of this investigation. The central lines could be determined therefore with sufficient accuracy by simple measurements on the records rather than by tedious calculation. This practice was followed for a large number of the records displayed in this paper.

The sets of superimposed records were each confined to convenient periods of about one solar rotation. This was adequate to establish a lower envelope and in addition gave the possibility of investigating longer-term changes in brightness distribution. Before a comparison could be made between different sets of records, however, further corrections had to be made to allow for annual effects.

The first annual effect is the changing declination of the Sun. The time of passage of the Sun through an aerial beam varies with the secant of the Sun's declination. This causes a change in the width of the record of the Sun's passage through the aerial beam, and this must be taken into account before a comparison is made between records separated in time by many days.

A second annual effect is the change in apparent size of the Sun during the year. Unlike the effect of declination changes, this does not affect the gain calibration procedure but it has simply the effect of increasing the size of the record in height and width. The changes are small, being less than 4 per cent.

A third annual effect is the change in the inclination of the Sun's axis relative to that of the Earth (and hence the aerial system) during the course of the year. This may result in changes in the recorded apparent width of the disk if this is non-circular.

III. Results

In Figure 1 are shown five sets of records obtained over a period of 9 months. A well-defined lower envelope is seen on these records and its position has been indicated for each set of observations. The areas of enhanced brightness are seen above a constant base level.

This lower envelope has the same height (within a few parts in 100) for all sets of observations. After the corrections have been applied it is found that the shapes of the curves are similar. Hence the results provide good evidence for the existence of a " quiet " component of the radio-frequency emission from the Sun.

It may be noted that the lower envelope is occasionally approached when there is little sunspot or other activity on the solar disk. We infer from this that during the present part of the solar cycle the emission sometimes falls to the base level at this wavelength, which corresponds to an apparent disk temperature of approximately 7×10^4 °K.

Although the distribution of brightness appears to remain constant during the period of the measurements, a careful comparison shows a very small systematic change in the apparent angular width of the Sun over this period. In Figure 2 the measured angular width of the solar disk, as derived from the five determinations of quiet-Sun brightness distribution, is plotted against the inclination of the Sun's axis (projected on the disk) relative to the direction on the disk of the strip which is being scanned at any instant by the aerial. Very little change in the apparent width of the solar disk is found as the direction of scanning changes. The small change that is present appears to indicate that the Sun is widest in the equatorial region. This latter conclusion was reached also by Blum, Denisse, and Steinberg (1952) who found that the solar disk had a markedly elliptical shape at metre wavelengths. Our results, however, are not conclusive, and further work over a much larger range of angles will be necessary before definite conclusions can be reached.

The effect found at 21 cm wavelength is small enough to justify the assumption of circular symmetry for the purposes of a preliminary analysis. On the basis of this assumption it is possible to determine the radial distribution of brightness over the disk. In Figure 3 is shown the radial brightness distribution over the solar disk as calculated from the five separate one-dimensional distributions of Figure 1 by means of an approximate method. It will be noted that the shapes of the curves are strikingly similar.

All of the diagrams shown in Figures 1, 2, and 3 are smoothed by the effect of the finite width of the aerial beam. Various methods are available to reduce this effect, although it is not possible to resolve features which have an angular width which is much less than that of the aerial beam. In Figure 4 (full line) the effect of the aerial has been partly removed from the curve of Figure 3 (c) by use of a simple iterative process described by Bolton and Westfold (1950). As a check, the effect of passing the aerial beam over the distribution of Figure 4 has been calculated and the one-dimensional brightness distribution found. This is shown in Figure 5 together with the experimental curve which was the starting point of the analysis.
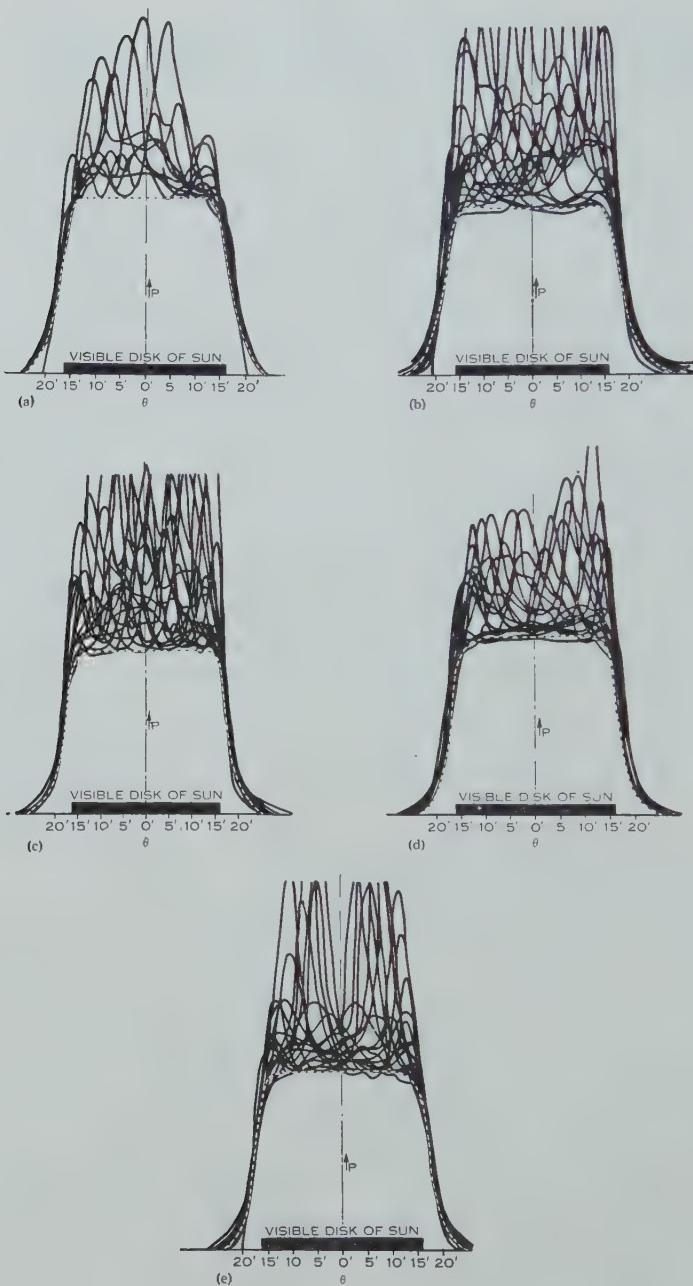
Fig. 1.—Superimposed daily records of the one-dimensional radio brightness distribution across the Sun during 1952.   (a) March 15 to April 9, (b) June 20 to July 10, (c) August 22 to September 12, (d) September 13 to October 3, (e) October 7 to October 28.   The dotted lines show the estimated base levels. P is the power received, in arbitrary units. θ is the angular separation between the centre of the aerial beam (scanning strip) and the centre of the optical disk of the Sun in minutes of arc.

The curve of Figure 4 is consistent with observations but these could equally well be satisfied if the bright ring round the edge of the solar disk were considerably sharper than is shown in Figure 4. The amount of energy in this bright ring can be calculated from this curve, but only the outer limit of its width can be derived. In Figure 4 a dotted line has been drawn to indicate one of an infinite number of distributions that could not be distinguished from that obtained from the measurements. Marked limb-brightening, of course, is a feature of all of these distributions.
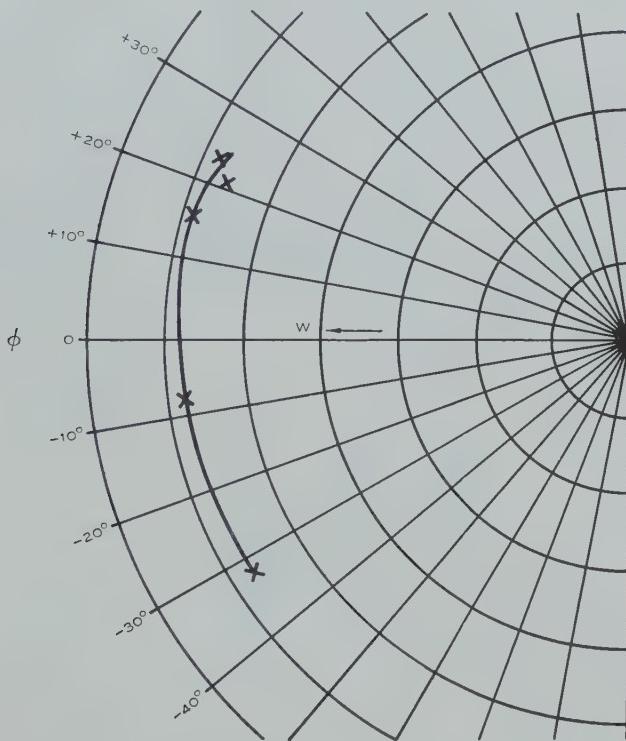
Fig. 2.—Apparent width W of the solar disk, between half-amplitude points on the curves of one-dimensional brightness distribution.
φ is the angle that the scanning strip on the disk makes with the projection on the disk of the Sun's axis.

From the brightness distribution at only one frequency it is not possible to derive the physical characteristics of the solar atmosphere. It is possible, however, to eliminate various possibilities by comparing the experimental curve with those derived theoretically from various solar models. A comparison has been made with some curves published by Smerd (1950) who assumed chromospheric and coronal densities derived from Cillié and Menzel (1935) and from Baumbach's formula as modified by Allen (1947) together with an isothermal corona and an isothermal chromosphere with several assumed values of electron
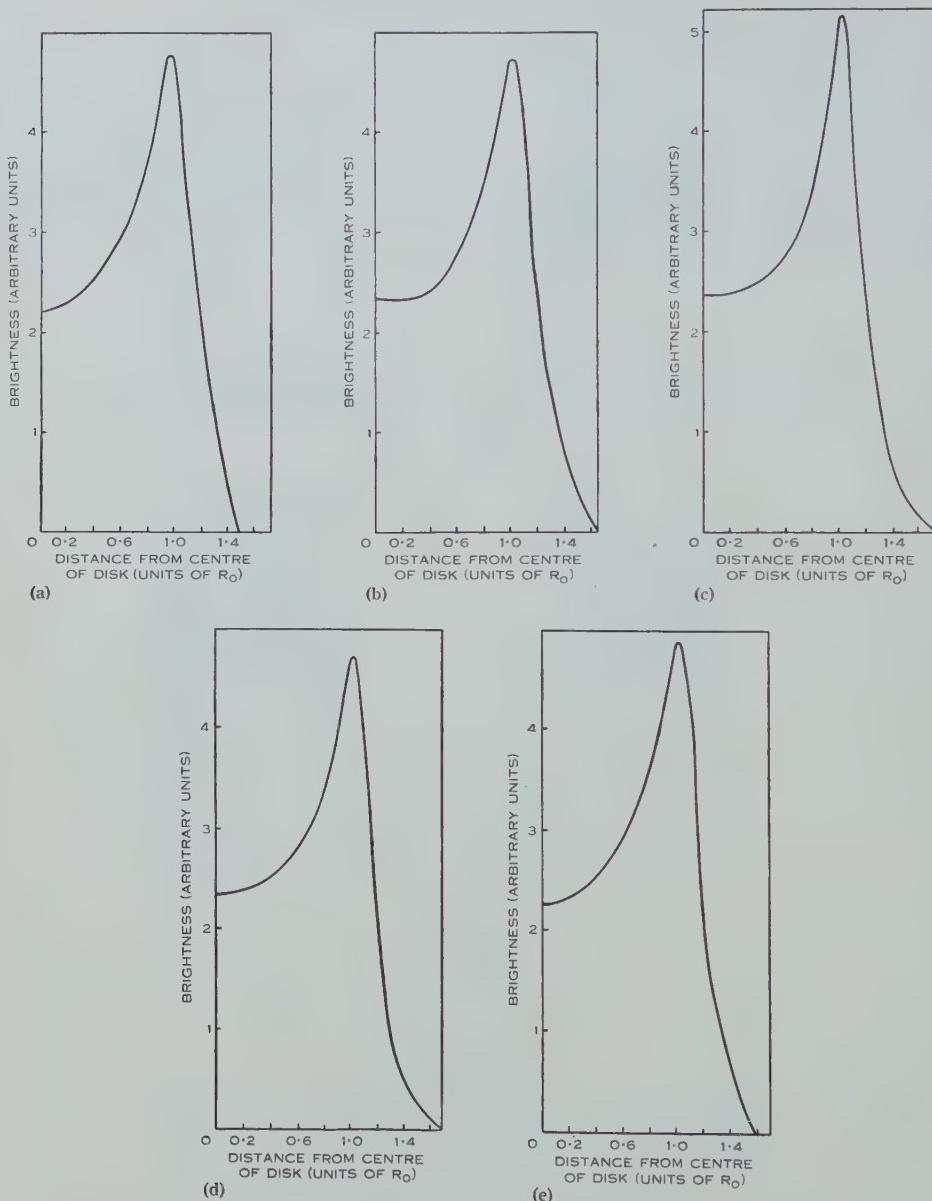
Fig. 3.—Radial distributions of radio brightness across the solar disk, derived from the one-dimensional distributions shown by dotted lines in Figure 1. R_0 is the radius of the visible disk.

temperatures. Smerd* (personal communication) has recently extended his calculations to include a greater range of temperatures and densities.

The curves published by Smerd for wavelengths close to 21 cm show the limb-brightening feature of the experimental distribution but show differences in

* The writers are indebted to Mr. Smerd for the use of these unpublished calculations.

detail. A comparison of the experimental distribution and the unpublished curves calculated for a wavelength of 21 cm reveals a reasonably good agreement, in the region from the centre of the Sun to a little beyond the optical limb, for a model in which the chromospheric temperature is 10^4 °K, the coronal temperature
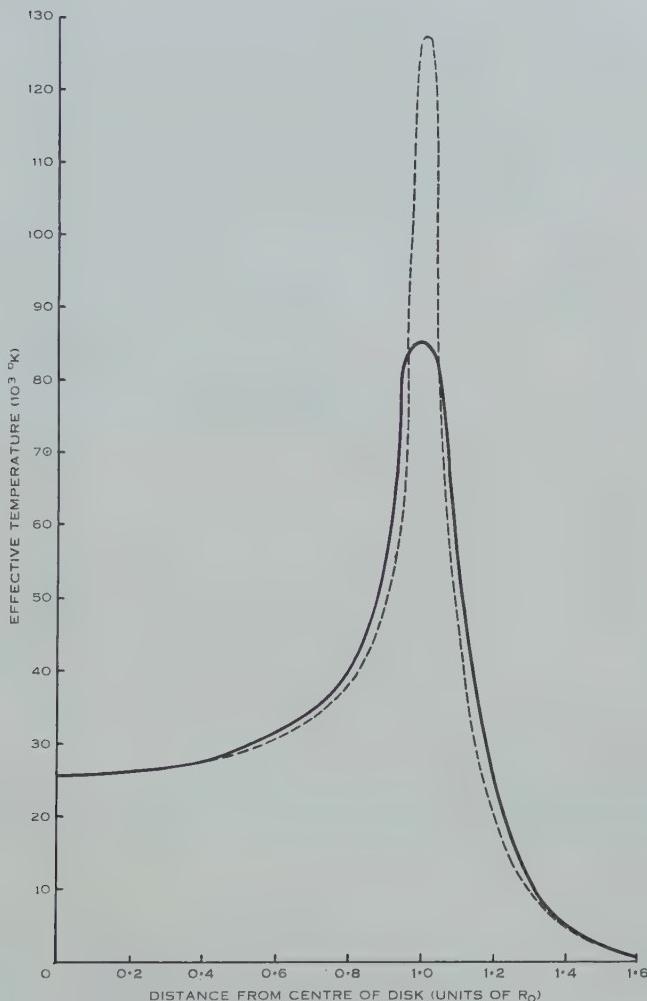
Fig. 4.—Radial brightness distribution across the Sun, derived from (c) of Figure 3. The effects of the finite beam width of the aerial have been partly removed (full line). The dotted line shows a typical distribution which would produce an effect on the aerial similar to that shown by the full line.

is 3×10^6 °K, and the coronal densities are as given by the Baumbach-Allen formula.

The agreement is still good if the assumed value of coronal temperature is reduced to 3×10^5 °K provided that the coronal densities are reduced to one-half of the values given by the above formula. Hence the experimental brightness

distribution at 21 cm does not lead to unique conclusions with respect to the coronal temperature.

In the region well outside the visible disk all the calculated distributions of brightness, for the ranges of coronal densities and temperatures quoted above, decrease more rapidly with increasing distance from the centre than does the observed distribution. This suggests that the actual rate of decrease in the outer corona is less than that given by the Baumbach-Allen formula.
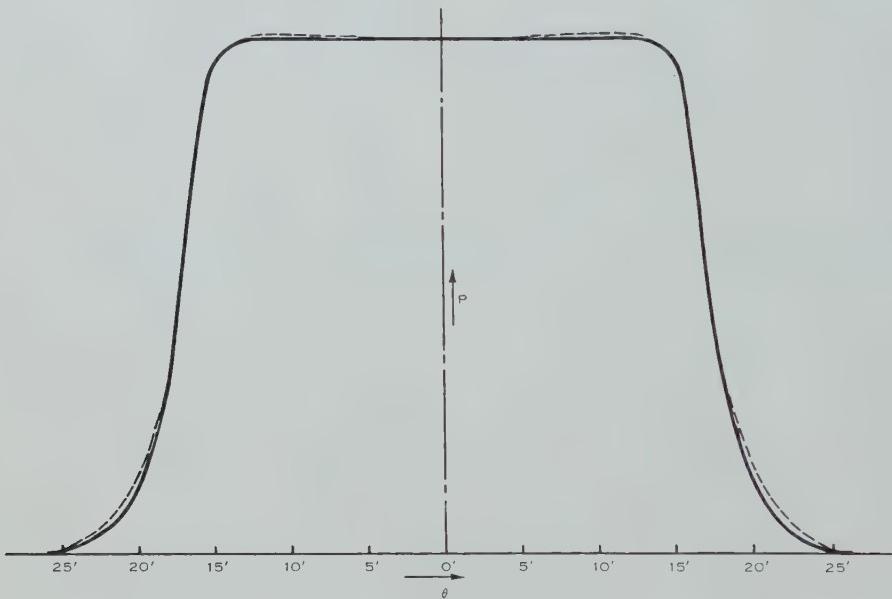
Fig. 5.—Check on Figure 4. The full line shows the experimental one-dimensional brightness distribution from which Figure 4 (full line) was derived. The dotted line shows the calculated effect of scanning the curve of Figure 4 with an aerial of 3′ of arc beam width.

IV. Conclusions

The presence of a clearly defined lower boundary to the daily curves of one-dimensional radio-brightness distribution for the Sun gives conclusive evidence for the existence of a base or quiet level in the radio-frequency emission from the Sun.

The solar disk appears to be nearly circularly symmetrical at a wavelength of 21 cm.

If circular symmetry is assumed, then the derived radial brightness distribution at a wavelength of 21 cm shows marked limb-brightening.

The distribution of radio brightness is in fair agreement with calculations from solar models involving a 10^4 °K chromosphere and a $0 \cdot 3$–$3 \cdot 0 \times 10^6$ °K corona, provided that coronal densities lie between $0 \cdot 5$ and $1 \cdot 0$ times the values usually quoted.

V. Acknowledgments

The authors wish to thank Dr. J. L. Pawsey and Mr. S. F. Smerd for helpful discussion of the work described in this paper.

VI. References

ALLEN, C. W. (1947).—*Mon. Not. R. Astr. Soc.* **107** : 426.

BLUM, E. J., DENISSE, J. F., and STEINBERG, J. L. (1952).—*Ann. Astrophys.* **15** : 184.

BOLTON, J. G., and WESTFOLD, K. C. (1950).—*Aust. J. Sci. Res.* A **3** : 19.

CHRISTIANSEN, W. N., and WARBURTON, J. A. (1953).—*Aust. J. Phys.* **6** : 190.

CILLIÉ, G., and MENZEL, D. H. (1935).—Circ. Harv. Astr. Obs. No. 410.

DENISSE, J. F. (1950).—*Ann. Astrophys.* **13** : 185.

VAN DE HULST, H. C. (1949).—*Nature* **163** : 24.

MARTYN, D. F. (1946).—*Nature* **158** : 632.

MARTYN, D. F. (1948).—*Proc. Roy. Soc.* A **193** : 44.

PAWSEY, J. L. (1946).—*Nature* **158** : 633.

PAWSEY, J. L., and YABSLEY, D. E. (1949).—*Aust. J. Sci. Res.* A **2** : 198.

REULE, A. (1952).—*Z. Naturf.* **7a** : 234.

SMERD, S. F. (1950).—*Aust. J. Sci. Res.* A **3** : 34.

STANIER, H. M. (1950).—*Nature* **165** : 354.

UNSÖLD, A. (1947).—*Naturwissenschaften* **34** : 194.

WALDMEIER, M., and MÜLLER, H. (1948).—Astr. Mitt. Zurich No. 155.

A HIGH–RESOLUTION AERIAL SYSTEM OF A NEW TYPE

By B. Y. Mills* and A. G. Little*

Summary

A method of constructing an aerial system of high resolution but small area and low cost is described. Its application to the production of narrow pencil beams at metre wavelengths for investigations in radio astronomy is discussed. A small-scale model has been constructed to test the principle.

I Introduction

Recent studies of cosmic radio-frequency radiation have shown that its brightness distribution over the sky is complex. Sources of an angular size less than about $\frac{1}{4}°$ have been known to exist for some years and, now, extended sources of considerably greater size have been observed (Bolton 1952 ; Mills 1952), some of which appear to merge with the general background radiation. Interferometric methods, which were so useful in the early days of radio astronomy, have encountered serious difficulties when used for observing such a complex distribution. It therefore appears desirable to rely mainly on the use of pencil beam aerials of high resolving power for future work, and to reserve the use of interferometric methods for special applications.

A study of the available information suggests that a beam width of the order of $1°$ or less is desirable for such a pencil beam. For an aerial of conventional form at metre wavelengths this beam width would require a prohibitively large and costly structure so that an alternative solution has been sought.

A satisfactory solution is possible because the number of randomly distributed discrete sources which can be individually detected at metre wavelengths with a large aerial is determined by the beam width (or the resolution) rather than the gain of the aerial. This follows from the fact that in these circumstances the number of discrete sources with intensities above the detectable threshold will normally greatly exceed the number which may be separately resolved. Advantage can be taken of this to construct an aerial system of high resolution but relatively low gain, that is, small effective area, which sacrifices very little of the usefulness of a conventional aerial but which can be made at a fraction of the cost.

Such a system can be constructed from two long aerials arranged in the form of a cross. At the wavelengths we are considering, these aerials preferably consist of arrays of dipoles. In Figure 1 this is shown schematically, together with an idealized diagram of the outline of the aerial beam produced when the

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

arrays comprising the two arms of the cross are connected together in the same phase. An aerial diagram of this form is unsatisfactory for, although there is an enhanced response in the central region where a source is received by both arrays, the total solid angle over which reception occurs is very large. However, advantage may be taken of the presence of signals in both aerials from the central region and their phase coherence for, if the arrays are now connected in antiphase, there will be no response from this central region, while the " spokes " of the diagram will be unaffected. If, therefore, the connections between the arms of the cross are switched rapidly between the two conditions, a source which is in the solid angle common to both beams will deliver a modulated signal at the switching frequency, while the signal from a source which is received by one aerial alone has no modulation imposed. After amplification and detection the modulated signal may be picked out by a phase-sensitive detector and used to
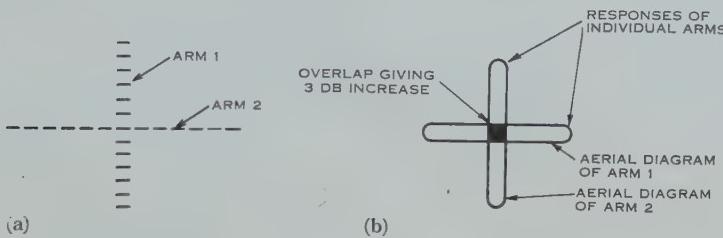
Fig. 1 (a).—Plan view of dipoles in cross arrangement.

Fig. 1 (b).—Idealized response of the cross arrangement, plan view.

deflect a pen recorder. The recorder then gives the integrated signal from within the central region so that, in effect, a pencil beam is produced which has a size determined by the *maximum* dimension of each array.

A similar method of using the common portion of overlapping beams to produce a narrower effective beam has already been used by one of the authors in an interferometer (Mills 1952).

II. ELEMENTARY THEORY

In order to appreciate the possibilities of an aerial of this type, it is necessary to examine its operation in more detail. Consider the response of the system to a signal of unit intensity originating in any arbitrary direction and, as a first approximation, assume that there is no interaction between the two arrays. We shall also assume that for each array the response in any direction may be fully represented by the voltage polar diagram, that is, the phase of the response is everywhere either equal or opposite to that in the direction normal to the array. This is true for aerials in which the current distribution is symmetrical about the mid point.

When the arrays are connected together in phase addition the received power is given by $P_1 = k(S_1 + S_2)^2$, where S_1 and S_2 are the amplitudes of the voltage polar diagrams of each aerial in the selected direction.

C

When the arrays are connected in phase opposition the received power is given by $P_0 = k(S_1 - S_2)^2$.

The modulated power delivered to the receiver is given by the difference of these quantities, that is,

$$P_m = 4kS_1S_2.$$

The composite power polar diagram is therefore given by the product of the voltage polar diagrams of each array. One consequence is that the recorder deflection will be negative where the voltage diagrams are of opposite sign. A further consequence is that trouble from side lobes will be accentuated, since in the planes in which one of the arrays has a wide angle of reception, the side lobes of the composite beam will be those appropriate to the voltage rather than the power diagram. For example, if the current distribution is uniform along each array, the first side lobes in the planes of the arrays are 20 per cent in amplitude instead of the normal 4 per cent. As any complexity in the composite diagram leads to great difficulty in the interpretation of the records, it is therefore imperative to minimize the side lobes.

A well-known method of eliminating side lobes in an aerial is to employ a current distribution across the aperture in the form of a normal error curve which produces a polar diagram of the same shape. Such a distribution has another advantage in the present case, for, consider two idealized polar diagrams

$$S_1 = e^{-k_1\theta^2}, \qquad S_2 = e^{-k_2\varphi^2},$$

where θ and φ are the two zenith angles in the planes of the arms, then the composite diagram is

$$S_c^2 = S_1S_2 = e^{-(k_1\theta^2 + k_2\varphi^2)}.$$

Thus a beam is produced without side lobes and of simple elliptical section everywhere. When $k_1 = k_2$ the beam section will be circular, which is the simplest possible case to deal with.

In practice a beam of exactly this shape cannot be produced because an infinite aperture is required. If, however, the aperture is finite and extended until the current has fallen to about 10 per cent., it is found that a sufficiently close approximation to an error curve is obtained. To obtain the same beam width this requires an aperture some 50 per cent. larger than that for a uniform distribution, but the advantages are obvious.

The above analysis of the operation is approximate only, for in general there will be an interaction between two arrays placed so close together. The effect of such an interaction may be obtained from a thermodynamical argument. Consider such an aerial in a constant temperature enclosure. No modulated output will then be produced, as the power received by the arrays will be a function of the enclosure temperature only and will be independent of any method of connecting them together. If the composite diagram formed from the cross product of the two voltage diagrams were to have an average value of zero this result would be given by the previous analysis. In general, however, the average value will not be zero as can be seen, for example, by integrating the idealized composite diagram above. The neglected interaction between the arrays must then be just sufficient to produce zero output. This point is considered in a practical example later.

From the thermodynamical argument it can be seen that the aerial, as described, measures only *differences* in brightness temperature between the complete reception angle of both arrays and the central solid angle common to both. However, a little consideration will show that the absolute value of the temperature over the central angle may be obtained by adding the temperature calculated from the amplitude of the modulated signal to the average temperature of the two arrays.

When the aerial is used in this way the results are similar to those which would be obtained from a conventional aerial of the same beam width with an attenuator connected between it and the receiver, the attenuation being roughly
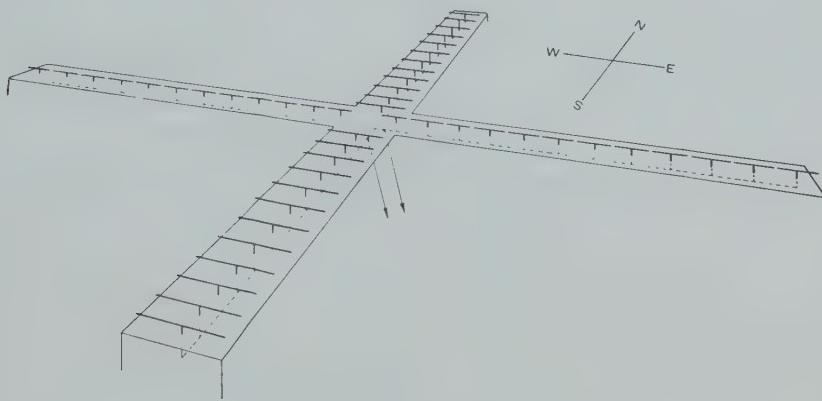
Fig. 2.—The experimental aerial system.

equal to the ratio of the area of the conventional aerial to that of the arrays forming the cross. When using a narrow pencil beam at wavelengths of a few metres for observing discrete sources, this attenuation is not very important because the resolution and not the sensitivity is the limiting factor. It does, however, lead to reduced sensitivity for observing the brightness temperature of extended distributions.

III. An Experimental Model

The construction of an aerial at metre wavelengths to produce a 1° beam, even when using the above method, is a large undertaking. It was decided, therefore, to construct first a small-scale model to test the principle and to allow experiments with possible designs. A sketch of this model is shown in Figure 2. It operates at a frequency of 97 Mc/s and the arms of the cross are 120 ft in length. They are arranged in the north-south and east-west directions. The beam width is 8°.

The east-west arm consists of a line of folded dipoles, end to end, backed by a wire-mesh reflecting screen. The dipoles are fed from a twin-wire transmission line stretching the length of the arm which is itself fed in the centre and terminated at each end by matching resistors. Quarter-wave resonant stubs are used to couple the dipoles to the feed line, the currents in the dipoles being

adjusted by changing the point at which they are connected to the stub as shown in Figure 3   Standing waves on the feed line are kept low by adding capacity at appropriate places.   The north-south arm consists of an array of full-wave dipoles similarly fed   The beam is swung in declination by changing the phases of the currents in the dipoles of this arm.   Phase changing is performed by changing the points of connection of the stubs to the feed line.

The currents are adjusted in each arm to coincide with the normal error curve which has a value at the ends of the arrays of about 10 per cent. of that at the centre.   Only about one-half of the power collected by an array is fed to the receiver, the remainder being dissipated in the matching resistors at the end of each feed line.   Each arm is connected to its own preamplifier, the outputs of which are combined through a phase switching arrangement as described above.   Further amplification, detection, and recording are accomplished in a conventional manner (Mills 1952).
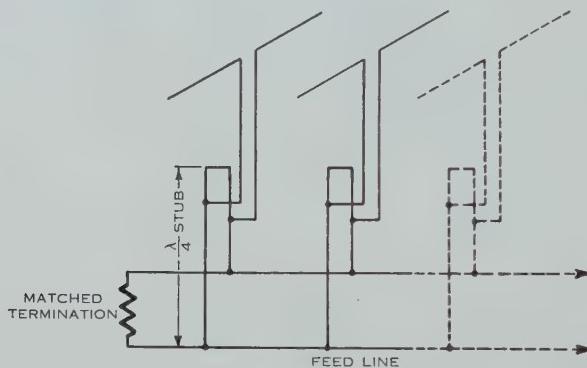
Fig. 3.—Arrangement of dipole feeds in experimental aerial.

The factors involved in observing uniform temperature distributions can be seen very clearly with this aerial.   In order to reduce complications due to cross coupling between the arrays the central dipoles of each were omitted. This results in relatively large distances between the closest dipoles so that cross coupling is small.   An attenuation of more than 40 db was measured between arrays.   The effect of omitting the central dipole, however, is to subtract its radiation field from the total, so that the voltage polar diagrams of each array, which would normally be zero outside the central beam, are now negative in that region.   Sources in the spokes of the diagram of Figure 1 therefore produce a negative deflexion.   It is easily shown that the average value of the composite diagram is now zero so that, if the aerial is pointed at a uniform temperature distribution, the positive deflexion over the central beam is counterbalanced by the negative deflexion produced over the much larger solid angle of the spokes of the diagram, and the net deflexion is zero.   If a different construction were utilized to eliminate this depression of the polar diagram, then, as was shown before, the cross coupling between aerials which is introduced produces a similar effect.

Sample records obtained with this experimental model are shown in Figures 4 and 5. Two quantities are recorded in each case, the output from the phase-sensitive detector in the upper graph and the receiver output level in the lower. The latter is a measure of the average temperature of the two arrays   The addition of the two temperatures derived from these graphs gives the actual temperature averaged over the central beam   Calibration marks are shown every 20 min when the preamplifiers are connected to cold resistors for a period of 1 min.
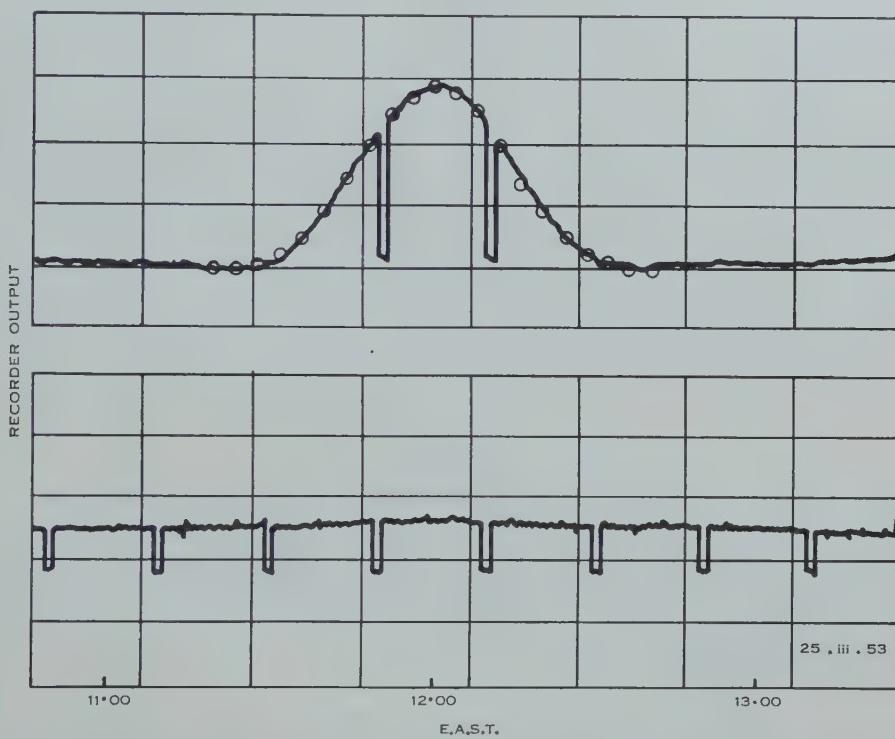
Fig. 4.—A record of the Sun showing the agreement between the observed and computed polar diagrams.

A record obtained on the Sun at a declination of $+1°$ is shown in Figure 4. The sensitivity was much reduced for observing such a strong source. Points derived from the computed polar diagram are shown superimposed on the record for comparison ; the agreement is excellent. The lower graph shows only a slight increase as the Sun passes through the aerial beams, the major part of the deflexion being due to the galactic background. The declination of $+1°$ represents an inclination of the beam of $35°$ from the vertical.

Figure 5 shows portion of a record obtained at a declination of $-34°$, that is, with the beam pointing vertically upwards. The comparatively strong " point " source 03–3 (Mills 1952) is shown first, then the weaker source 04–3. The effect of an " extended " source is illustrated when the Galaxy at about

galactic longitude 220° crosses the beam. The record illustrates one advantage of this system, for the sky temperature is divided into a slowly varying background which is recorded at relatively low sensitivity on the bottom graph and a detailed structure which is recorded at high sensitivity on the upper. A conventional aerial operating at the same sensitivity would require some system of backing off the output to keep the deflexion within the range of the recorder.

A survey of the southern sky is now in progress and an analysis of this and similar records will be given in a subsequent paper. It is interesting to note, however, that the resolution of the model aerial has been sufficient to detect radiation from the large Magellan Cloud and to show that the centre of the Galaxy is narrower than previously suspected from earlier low-resolution surveys.
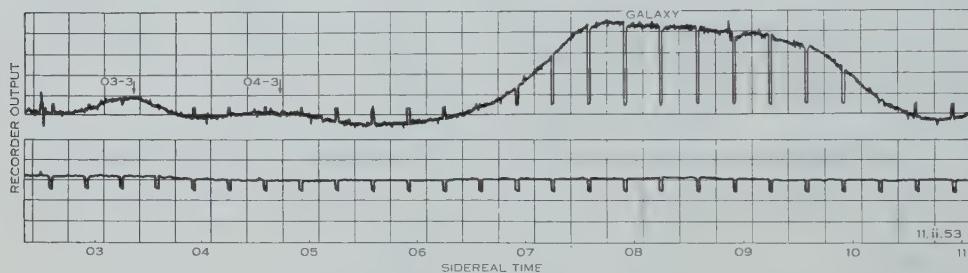
Fig. 5.—A record centred on declination —34°, showing two discrete sources and the Galaxy at longitude 220°.

The experimental work which has been performed with the small-scale model has shown that the principle of operation is sound and has demonstrated the feasibility of a full-size aerial. Work is in progress on the design and construction of an aerial which will operate at a frequency of about 80 Mc/s and have a beam width of less than 1°. It will be capable of surveying approximately one-half of the celestial sphere.

IV. ACKNOWLEDGMENTS

The authors are indebted to Dr. J. L. Pawsey for valuable discussions on the many problems encountered in this work and to Mr. K. Sheridan for his able assistance with the technical details of the aerial.

V. REFERENCES

BOLTON, J. G. (1952).—Extended sources of galactic noise. U.R.S.I. Report 1952.
MILLS, B. Y. (1952).—*Aust. J. Sci. Res.* A **5** : 266-87.

BUOYANT MOTION IN A TURBULENT ENVIRONMENT

By C. H. B. Priestley*

Summary

Solutions are given of the simultaneous equations for the vertical velocity and temperature of an element of fluid moving under buoyancy and subject to continuous mixing of heat and momentum with its environment. Three distinct modes of behaviour result : (*A*) ascent followed by damped oscillations, (*B*) asymptotic ascent to an equilibrium level, (*C*) *absolute buoyancy* in which the ascent rate increases indefinitely. For an environment in which the lapse rate is subadiabatic the motion is of type *A* for sufficiently large elements but may become *B* for the smaller elements ; in super-adiabatic lapse rates the mode is *C* for sufficiently large elements, and *B* for the smaller elements, which are in no way unstable. The *mode* of motion is independent of the initial conditions but the scale of the motion is not.

The same formulation applies within known limits to the ascent of saturated air, and applications made to atmospheric convection include the verification of a formula for the period of large cloud tops oscillating in a stable layer.

I. INTRODUCTION AND SCOPE

The motions which might ensue, and the stability of these motions, when a fluid is heated from below was first discussed in a formal manner by Rayleigh (1916). This classical paper, and those that followed it, require no summary here ; one has recently been provided by Sutton (1950), who goes on to offer an explanation for another mode of motion observed under laboratory conditions and described as *columnar*, in contrast to the regular cellular pattern which had been the subject of the earlier work. It has been demonstrated that motion will not necessarily ensue in either case when the layer of fluid is thin, even though its stratification is favourable, owing to the prohibiting effects of viscosity and molecular conduction. The thinner the layer, the more " unstable " the stratification can be without motion occurring.

Valuable as these studies are, they are not entirely suited to the treatment of the phenomenon as it occurs in nature. Their shortcomings in this context are the assumptions that the heating at the bottom is uniform and that the only buoyant forces which operate within the medium are those which arise directly as a result of this heating or by advection from the motion which the heating sets up. Natural surfaces are subject to both small- and large-scale irregularities in their thermal properties ; there are similar irregularities in their radiative properties so that in the atmosphere, bounded by a natural surface which emits and receives radiation, the scope for variety is great. The irregularities are passed on in the form of temperature contrasts, and in other ways, to the layers

* Section of Meteorological Physics, C.S.I.R.O., Melbourne.

of the medium most closely in contact with the surface. Moreover, when the medium is already in turbulent motion there will in general be a continuous source of creation of irregular buoyancy within the medium itself ; when an element has been displaced vertically it will in general find itself at a different density from its environment and will henceforward be subject to buoyant forces as well as those inherent in the turbulence until equilibrium is re-established.

Whereas Rayleigh and his followers, by assuming a regular patterned structure, were able to formulate equations which took account of the condition of every element of the medium, this is no longer possible in the natural phenomenon where the buoyant elements are arranged in space, and to some extent also in time, in a disorderly manner. For the latter we require in the first place a treatment of the changes in motion and temperature experienced by an individual buoyant element, preferably in terms which permit of statistical recombination at a later stage. In the normal phraseology of meteorology the approach is along the lines of the *parcel* as opposed to the *slice* method (Normand 1946 ; Petterssen *et al.* 1946). But the parcel method does not appear to have been extended in any satisfactory manner to allow for the simultaneous mixing, both of heat and momentum, which must take place between the element and its environment according to the laws of turbulent exchange.

An allowance for these processes, and for the consequent need to solve two simultaneous equations instead of one, forms the essential contribution of this paper. This leads to a formulation of single-cell convection which embraces some aspects of the process of *entrainment* of air, though in a manner different from that adopted by other workers (see Austin (1951) for a summary). The present treatment is not confined to the " unstable " stratification since the origins of elemental buoyancy are present in nature even when the medium is cooled from below or a " stable " stratification is otherwise brought about.

For mathematical ease and resultant physical clarity, the assumption has been made that the element is small, in the sense that the mixing of its excess heat and vertical momentum does not affect the average condition of the environment. It is not thought that this assumption is fundamentally restrictive—in other studies it has been found that the slice method modifies the results obtained from the parcel method but does not change their character.

II. The General Equations

The problem to be solved is that of describing the motion and temperature behaviour of an element of fluid moving under its own buoyancy and subject to the turbulent transfer of heat and momentum into an environment which is at rest and remains in a steady state. The treatment will first be restricted to a medium of uniform constitution although, as indicated in Section IV, the ascent of saturated air may be brought within the framework of the same equations and solutions.

It will be assumed that the relative variations in density in space and time are slight as compared with those in velocity, so that density becomes important only in so far as it affects the buoyancy (Rayleigh loc. cit.). If then T and T_e

represent the temperature of the element and environment respectively, w the upwards velocity, and Γ the lapse rate of temperature with height experienced by an element ascending adiabatically, the equations are

$$\frac{\partial w}{\partial t} + u\frac{\partial w}{\partial x} + v\frac{\partial w}{\partial y} + w\frac{\partial w}{\partial z} = \frac{g}{T_e}(T - T_e),$$

$$\frac{\partial T}{\partial t} + u\frac{\partial T}{\partial x} + v\frac{\partial T}{\partial y} + w\frac{\partial T}{\partial z} = -w\Gamma,$$

the effects of molecular viscosity and conductivity being neglected from the start since they are much smaller than others which will appear when the equations are transformed.

It is required to apply these equations to an aggregate of elements which have a common measure of velocity, $\bar{u}$, $\bar{v}$, $\bar{w}$, and of temperature $\bar{T}$. Writing in the usual fashion $u = \bar{u} + u'$ etc., and using the equation of continuity, the equations become

$$\dot{\bar{w}} = \frac{g}{T_e}(\bar{T} - T_e) - \overline{\frac{\partial}{\partial x}(u'w') + \frac{\partial}{\partial y}(v'w') + \frac{\partial}{\partial z}(w'w')},$$

$$\dot{\bar{T}} = -\bar{w}\Gamma - \overline{\frac{\partial}{\partial x}(u'T') + \frac{\partial}{\partial y}(v'T') + \frac{\partial}{\partial z}(w'T')},$$

where the dot denotes the time derivative following the mean motion. Henceforward we shall reserve the term *elements* for the aggregates so defined and *sub-elemental* when reference is made to finer structure. The motion of the element is defined as the mean of that of the composing particles, so that the element is regarded as composed of a changing set of particles and some aspects of entrainment are thereby formally taken into account. It will be assumed that the buoyancy occurs in a medium already in turbulent motion; that turbulent transfer coefficients for momentum and heat, K_1 and K_2, exist and are independent of position on the sub-elemental scale, that they are determined by processes other than the buoyant motion and so are independent of $\bar{w}$. The equations may then be written

$$\dot{\bar{w}} = \frac{g}{T_e}(\bar{T} - T_e) + K_1\overline{\nabla^2 w}, \quad \ldots\ldots\ldots\ldots \quad (1)$$

$$\dot{\bar{T}} = -\bar{w}\Gamma + K_2\overline{\nabla^2 T}. \quad \ldots\ldots\ldots\ldots\ldots \quad (2)$$

These are to be applied in what follows to elements of constant finite size within and around which there is some overall pattern in w and T apart from the random variations of the turbulence. While it is part of the complete problem to determine just what this pattern will be, this aspect will not be investigated here where the concern will be with broader features of the behaviour, and we may write

$$K_1\overline{\nabla^2 w} = -\frac{c_1 K_1}{R^2}\bar{w},$$

$$K_2\overline{\nabla^2 T} = -\frac{c_2 K_2}{R^2}(\bar{T} - T_e),$$

where R is taken to characterize the size (radius) of the element and c_1 and c_2 are numerical factors which depend on the exact form of the patterns. Defining then

$$k_1 = \frac{c_1 K_1}{R^2}, \quad k_2 = \frac{c_2 K_2}{R^2}, \quad \dots\dots\dots\dots\dots \text{(3)}$$

the equations of the element are written as

$$\dot{w} = \frac{g}{T_e}(T - T_e) - k_1 w, \quad \dots\dots\dots\dots\dots \text{(4)}$$

$$\dot{T} = -w\Gamma - k_2(T - T_e), \quad \dots\dots\dots\dots \text{(5)}$$

the bars now being superfluous.

It might have been acceptable to write down equations (4) and (5) without explanation, as affording a satisfactory starting point for treating the problem in hand. This has indeed been done with equation (4) (e.g. Scorer and Ludlam 1953), though apparently not with the two equations together. The derivation indicated above has the advantages of showing that their application is to elements of constant size and of providing a proper background for the physical interpretation of k_1 and k_2. As in most applications of turbulence theory, intuitive or dimensional arguments have replaced rigorous mechanistic ones at some points; in particular the distinction between turbulent components w' and those deriving from the buoyancy is artificial and there must remain some possibility, which is excluded from what follows, that k_1 and k_2 are not strictly independent of w.

k_1 and k_2 will be referred to as *mixing rates* for momentum and heat respectively. The mixing rate will depend on the ratio of surface to volume of the element, and one might intuitively expect that it will be greatest for the smallest elements and vice versa. The issue is not quite so straightforward, since an underlying principle of modern ideas on turbulence is that K is itself a function of R, increasing as R increases: but the strong indication, both from theory at the smaller scales (Weizsäcker 1948) and from empiricism up to the largest scales (Richardson 1926), is that the rate of increase is approximately as $R^{4/3}$, whence from (3) the intuitive expectation is confirmed subject to the constancy of c_1 and c_2. k_1 and k_2, then, will depend on the general level of turbulence in the environment; in particular this will prevent them from becoming discordant in magnitude or, in other words, they may tend to very large or very small values together, but never separately. Apart from this broad control, it is the dependence of k on R which leads to the more interesting novel features in the interpretations of the solutions which follow, and will accordingly be stressed. The general level of turbulence will be treated as given, and k_1 and k_2 regarded as constants during the life history of a given element; distinguishing features of the solutions at different values of k will be interpreted primarily as distinctions between modes of behaviour of elements of different size.

The form factors c_1 and c_2 depend only on the *shape* of the distribution of w and T within the element, not on the intensity. Whether they remain constant, to an order of magnitude, over the lifetime of a given element and

over a wide range of sizes cannot be stated until more is known of the character-
istics of turbulence at sub-elemental scales : the constancy is here assumed as a
reasonable working hypothesis, of a type related to the similarity assumptions
common in modern turbulence theory. Some values are given in Appendix I.

It remains to solve (4) and (5) with k_1 and k_2 constant, under the condition
for a steady environment

$$\dot{T}_e = w \frac{\partial T_e}{\partial z}. \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots (6)$$

Writing $T' = T - T_e$, we obtain from (4) and from (5) and (6)

$$\dot{w} = \frac{g}{T_e} T' - k_1 w, \quad \dots\dots\dots\dots\dots\dots\dots (7)$$

$$\dot{T}' = -w\left(\frac{\partial T_e}{\partial z} + \Gamma\right) - k_2 T', \quad \dots\dots\dots\dots\dots (8)$$

from which T' can be finally eliminated, yielding

$$\ddot{w} + (k_1 + k_2)\dot{w} + \left[\frac{g}{T_e}\left(\frac{\partial T_e}{\partial z} + \Gamma\right) + k_1 k_2\right]w + \frac{1}{T_e}\frac{\partial T_e}{\partial z}w(\dot{w} + k_1 w) = 0.$$
$$\dots\dots\dots\dots\dots\dots\dots (9)$$

This is to be regarded primarily as the general equation for the buoyant
motion of a single element, with assigned values for k_1 and k_2 and known environ-
mental conditions. Having solved for w, the corresponding solution for T'/T_e
is obtained from (7). An alternative viewpoint would be that, if the behaviour
of elements in a known environment can be observed in detail, equation (9)
becomes an equation providing evidence about the turbulence factors k_1 and k_2
and their dependence on scale.

For ease in what follows, auxiliary " rates " $\varkappa$, λ, μ are defined by

$$\varkappa^2 = \frac{g}{T_e}\left(\frac{\partial T_e}{\partial z} + \Gamma\right) + k_1 k_2,$$

$$\lambda^2 = \left|\frac{g}{T_e}\left(\frac{\partial T_e}{\partial z} + \Gamma\right)\right|,$$

$$\mu^2 = \left|\frac{g}{T_e}\left(\frac{\partial T_e}{\partial z} + \Gamma\right) - \frac{(k_1 - k_2)^2}{4}\right| = \left|\varkappa^2 - \frac{(k_1 + k_2)^2}{4}\right|.$$

III. PROPERTIES OF THE BASIC SOLUTIONS

While equation (9) can be solved numerically under any known conditions,
it is more enlightening to derive analytical solutions under particular environ-
mental conditions which make this possible.

In nature it is observed that the temperature variations at a given level are
at all times small as compared with the absolute temperature. Close above a
factory chimney or open fire this will not be so, but excluding these cases it may
be seen from (7) that the last term in (9) is

$$\frac{g}{T_e}\frac{\partial T_e}{\partial z}w \times \frac{T'}{T_e},$$

in which w is multiplied by a quantity which remains small. So long then as $\varkappa^2$, the other multiplier of w, is not small the last term in (**9**) may be neglected. It must be stressed that this simplification is made without putting any restriction on the value of w itself, so that large velocities are not excluded.

It then becomes evident that the type of motion will depend on the multipliers of the remaining terms, that is, on $\varkappa^2$ and (k_1+k_2). Although in practice $\varkappa$ may vary with height, the chief characteristics of the possible modes of behaviour may be most clearly brought out by solving the equation for $\varkappa$ constant ; simple analytical solutions are then obtained which can be recombined when practical application so requires.   Natural convection in a layer of constant $\varkappa$, which for practical purposes can be identified as one of constant lapse rate, is therefore governed by the second-order equation with constant coefficients,

$$\ddot{w}+(k_1+k_2)\dot{w}+\varkappa^2w=0, \quad \ldots\ldots\ldots\ldots\ldots (\mathbf{10})$$

the form of whose solutions is familiar. The discussion of them will be eased by reference to Figures 1 and 2 ; the first represents the special case $k_1=k_2$ but the second is typical of the more general conditions where $k_1 \neq k_2$.
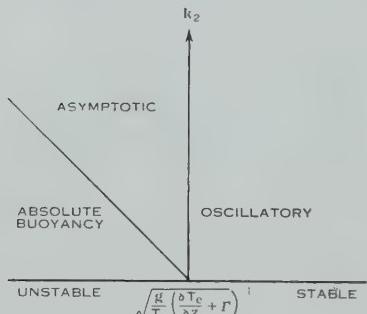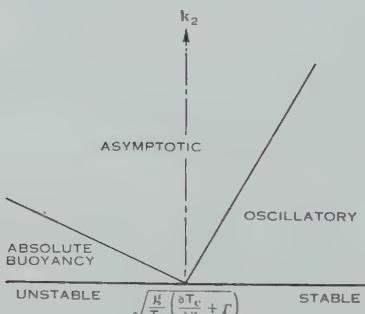
Fig. 1   Fig. 2

Fig. 1.—Modes of motion as a function of lapse rate and mixing rates for $k_1=k_2$.
Fig. 2.—Modes of motion as a function of lapse rate and temperature mixing rate k_2 for $k_1=2k_2$.

(i) When

$$\frac{g}{T_e}\left(\frac{\partial T_e}{\partial z}+\Gamma\right)>\left(\frac{k_1-k_2}{2}\right)^2,$$

the solutions of (**10**) and (**7**) are given by

$$w=Ae^{-\frac{k_1+k_2}{2}t} \sin (\mu t+\varepsilon_1), \quad \ldots\ldots\ldots\ldots (\mathbf{11})$$

$$\frac{T'}{T_e}=\frac{A\lambda}{g}e^{-\frac{k_1+k_2}{2}t} \sin (\mu t+\varepsilon_1+\varepsilon_2), \quad \ldots\ldots\ldots (\mathbf{12})$$

where A and ε_1 are determined by the initial conditions and $\tan \varepsilon_2=2\mu/(k_1-k_2)$. This mode of motion will be called *oscillatory*, the element transcending an equilibrium level about which it subsequently executes damped harmonic oscillations. This solution can obtain under inversion, isothermal, or sub-adiabatic lapse rates ; if $k_1=k_2$, it holds for all elements under these conditions,

but if $k_1 \neq k_2$, only for sufficiently large ones (small k). For the very largest elements ($k \to 0$) the motion becomes strictly periodic with period $2\pi/\lambda$. This formula for the period in the extreme case has been given by Brunt (1927), but it does not appear to have been applied except as a possible explanation of microbarograph oscillations.

(ii) When

$$\frac{g}{T_e}\left(\frac{\partial T_e}{\partial z}+\Gamma\right)=\left(\frac{k_1-k_2}{2}\right)^2,$$

there results a motion which approaches its equilibrium level asymptotically, either from below or from above after a single overshooting according to the initial conditions. This is a special case of restricted interest and will not be discussed further.

(iii) When

$$\frac{g}{T_e}\left(\frac{\partial T_e}{\partial z}+\Gamma\right)<\left(\frac{k_1-k_2}{2}\right)^2,$$

the solutions are

$$w=A e^{-\frac{k_1+k_2}{2}t}\ \sinh\ (\mu t+\varepsilon_1),\qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (13)$$

$$\frac{T'}{T_e}=\frac{A}{g}e^{-\frac{k_1+k_2}{2}t}\left\{\frac{k_1-k_2}{2}\ \sinh\ (\mu t+\varepsilon_1)+\mu\ \cosh\ (\mu t+\varepsilon_1)\right\}.\ \ \ldots\ (14)$$

These represent two quite different modes of behaviour according as the exponential or the hyperbolic terms ultimately become dominant.

When $\varkappa^2$ is positive, the resulting mode will be called *asymptotic*; the exponential term is finally dominant and the element approaches its equilibrium level asymptotically without overshooting. For $k_1=k_2$ this mode is confined to the smaller elements in superadiabatic lapse rates, but for $k_1 \neq k_2$ it applies to sufficiently small elements under any conditions of lapse rate.   It is of interest to note that in both asymptotic and oscillatory motion T'/T_e remains small if small initially, so that these solutions are solutions of the complete equation (9) as well as of (10).

When $\varkappa^2$ is negative, that is, for sufficiently large elements in superadiabatic lapse rates, (13) and (14) give expressions for the velocity and buoyancy which ultimately increase exponentially with time.   This condition will be referred to as one of *absolute buoyancy*.   It is clear that (13) will ultimately fail to satisfy equation (9) but it may readily be seen that the effect of the last term in (9), which has been neglected, will be to enhance rather than suppress the acceleration.   The condition of absolute buoyancy is therefore real, the criterion for its realization being

$$-\frac{g}{T_e}\left(\frac{\partial T_e}{\partial z}+\Gamma\right)\geqslant k_1 k_2.\ \ \ldots\ldots\ldots\ldots\ldots\ldots (15)$$

Significance of Element Size

Apart from the recognition of the three characteristic types of motion, the important result of the foregoing analysis is that the mode of behaviour is determined entirely by the values of λ, k_1, and k_2. Physically this means that the type of motion depends on the environmental conditions (lapse rate and

general level of turbulence) *and on the size of the element*, and is independent of the initial conditions of velocity and buoyancy.

Buoyancy Length and Static Instability

As opposed to the *type* of motion, the *scale* of the motion does depend on the initial conditions. It is most conveniently characterized by the distance of the final equilibrium above the starting level which, following Priestley and Swinbank (1947), is called the *buoyancy length*; this concept applies in all instances of oscillatory and asymptotic motion and, since then w and $\dot{w}$ tend to zero as t tends to infinity, the length is obtained by direct integration of (**10**) as

$$L = \int_0^\infty w\,\mathrm{d}t = \frac{1}{\varkappa^2}[\dot{w}_0 + (k_1 + k_2)w_0],$$

whence from (**7**)

$$L = \frac{1}{\varkappa^2}\left[\frac{g}{T_e}T'_0 + k_2 w_0\right], \quad \dots\dots\dots\dots\dots\dots\dots \quad (\mathbf{16})$$

the suffix zero denoting initial values. For motion starting from rest

$$L = T'_0 \Big/ \left(\frac{\partial T_e}{\partial z} + \Gamma + \frac{k_1 k_2 T_e}{g}\right). \quad \dots\dots\dots\dots\dots \quad (\mathbf{17})$$

It is seen from (**16**) that an element given an infinitesimal impulse or temperature disturbance will come to rest after an infinitesimal distance, unless it is absolutely buoyant. The idea that elements are statically unstable when in a superadiabatic lapse rate is therefore true only in a limited sense. In order to create a finite motion the disturbance must be of finite intensity, or alternatively an infinitesimal disturbance must be applied to an element of sufficiently large size. This result is the counterpart of the classical one of Rayleigh which was referred to in the opening paragraph.

Critical Size for Absolute Buoyancy

The theory provides a criterion for the realization of absolute buoyancy in the form of a condition relating the mixing and lapse rates. At the present stage only rather crude empiricism can transform this into a condition for the critical size of element, but this is worth while even though it can hope to indicate no more than the order of magnitude involved. Since k_1 and k_2 are of the same order of magnitude we shall equate them in the present context, whence from $K = aR^{4/3}$ and equations (**3**) and (**15**) is obtained

$$R = (ca)^{3/2}\left\{-\frac{g}{T_e}\left(\frac{\partial T_e}{\partial z} + \Gamma\right)\right\}^{-3/4}. \quad \dots\dots\dots\dots \quad (\mathbf{18})$$

Richardson (1926) gives $a = 0 \cdot 2$ c.g.s. units and taking $c = 8$ (see Appendix I) Table 1 results.

TABLE 1

CRITICAL ELEMENT SIZE

$-(\partial T_e/\partial z + \Gamma)$ (°C/cm) ..	10^{-7}	10^{-6}	10^{-5}	10^{-4}	10^{-3}	10^{-2}
R (m)	1500	250	40	8	$1\frac{1}{2}$	$\frac{1}{4}$

In the free atmosphere, if the lapse rate exceeds the adiabatic by only a small fraction of the latter's value (10^{-4} °C/cm), the critical size is of the order of hundreds of metres, whereas in strong lapse rates very close to the ground it is of order of a metre. The sizes thus obtained support the application of the different modes of solution both to atmospheric convection and to the motion of, and transport of heat by, small elements close to the ground. The first application will be discussed in Section IV and the last, which involves a statistical recombination of the solutions, will be given in a later paper.

IV. APPLICATION TO ATMOSPHERIC CONVECTION

The scope for application of the foregoing to the problem of single-cell convection in the atmosphere is clear enough, but there remain a few points which merit some further discussion.

Dry and Saturated Ascent

The basic equations were derived for a fluid of uniform constitution and it is necessary to state to what extent they remain valid for the mixture of air, water vapour, and water drops which occurs in the atmosphere. So long as the air is not saturated the effect of the presence of water vapour on the specific heat and density is very slight and (**1**) and (**2**) are valid. Γ may then be identified as the dry adiabatic lapse rate Γ_d.

When the air becomes saturated (**2**) and (**5**) no longer hold because of the heat involved in changes in the water phase. The consequences of this may be represented by the same set of equations and solutions provided Γ is taken as some value between Γ_d and Γ_s, the saturated adiabatic rate, and equations (**3**) are modified to

$$k_1 = \frac{c_1 K_1}{R^2}, \qquad k_2 = \frac{\Gamma}{\Gamma_d} \cdot \frac{c_2 K_2}{R^2}. \qquad \ldots \ldots \ldots \text{ (19)}$$

Γ is dependent, within its known limits, on the rate of mixing between element and environment, but it cannot strictly be regarded as constant during the lifetime of an individual element. Nor, therefore, can k_2. But for the largest elements the mixing becomes unimportant, k_1 and k_2 tend to zero, and Γ becomes equal to Γ_s, and the behaviour is strictly determinate.

In general the formulation is not complete because Γ has not been exactly specified. A complete treatment would involve appropriate modification of (**5**) and addition of a third simultaneous equation to express the mixing of water. More complicated motions will then ensue. For example, a cloud element in asymptotic motion, when near its equilibrium level as at present defined, is likely if anything to exceed the environment in moisture content and so will evaporate some of its water and fall back towards a lower level.

Generation of Convective Motion

The vertical velocities which occur in association with clouds and dry thermals are considerably larger than those which are otherwise present and it follows that these are initiated by absolute buoyancy since this is the only mode of motion capable of magnifying an initial disturbance by an order of magnitude.

Table 1, in indicating critical sizes which are rather smaller than the dimensions of most clouds, would suggest the same conclusion.

The question then arises as to how buoyant elements of sufficient size may be generated if external disturbances on this scale are absent. Attention has recently been drawn by Scorer and Ludlam (1953) to some interesting consequences of wake formation, one of which, when considered against the background of the present solutions, suggests an appropriate mechanism. An element so small that its motion is bounded (asymptotic) will leave a residual wake of greater size, though of smaller intensity, than the element itself. On the present theory it is size rather than intensity which is critical and, given a superadiabatic lapse rate, repetitions of this process should eventually produce an element large enough to be absolutely buoyant, provided only that the wind shear is not so strong as to disrupt the coherence of the successive wake elements.
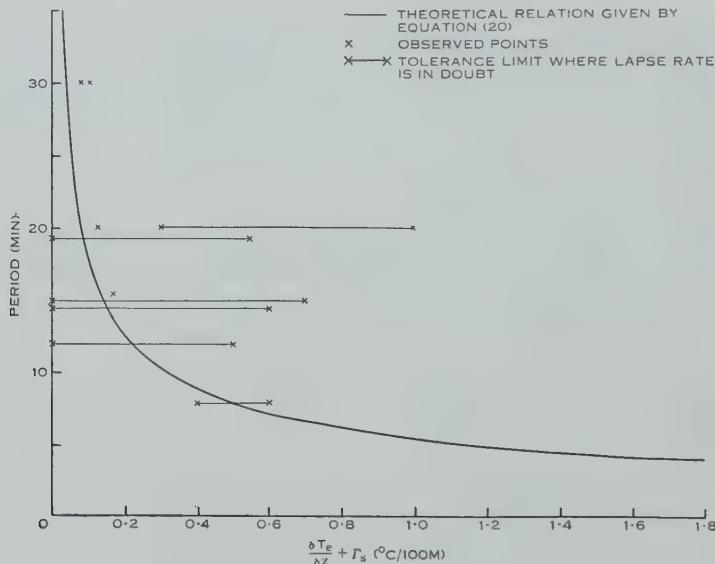
Fig. 3.—Period of cloud oscillation $v.$ lapse rate.

Final Stages

Generated by absolute buoyancy, the motion will eventually die away when a more stable layer is reached. Figures 1 and 2 show that, if all the variables are continuous, the motion cannot pass from absolute buoyancy to the oscillatory mode without passing through the asymptotic mode. Depending on the characteristics of the upper layer, the element will be brought to rest asymptotically or may penetrate further and become subject to oscillations; both types of final motion are observed in practice. The oscillatory mode has received little attention, but it occurs by no means rarely (Workman and Reynolds 1949; Smith 1951). For sufficiently large clouds the motion will then approach a simple harmonic oscillation of amplitude L and period

$$2\pi \Big/ \sqrt{\frac{g}{T_e}\left(\frac{\partial T_e}{\partial z}+\Gamma\right)}. \qquad \ldots\ldots\ldots\ldots (20)$$

Measurements of the periods of oscillating cloud tops have been made from aircraft operating near Sydney, and may be checked against lapse rates shown by the radiosonde ascent from Rathmines on the same day. The results are shown in Figure 3; in view of the type of data employed the agreement is regarded as good, bearing in mind that (**20**) represents a prediction of the order of magnitude of the period as well as of its variation with lapse rate.

On some occasions the radiosonde showed $\partial T_e/\partial z + \Gamma_s$ changing abruptly with height from a zero or negative to a markedly positive value, the levels of the stable layer being close to those of the reported oscillations. These are the conditions most favourable for large oscillations, as was confirmed by the reports. There was little difficulty on these occasions in identifying the appropriate value of $\partial T_e/\partial z + \Gamma_s$ but in others, where the change was more gradual, this could be estimated only within rather wide tolerance limits. Differences in time and place (not exceeding a few hours and 100 miles) add to the uncertainty.

A final point of quantitative appeal is provided by the observed amplitude of the oscillations. Taking average values from the 10 occasions, the amplitude of 800 ft with $\partial T_e/\partial z + \Gamma_s = +0 \cdot 25 \times 10^{-4}$ °C/cm yields $T_0' = 0 \cdot 6$ °C in (**17**); such a value is in harmony with other evidence (Byers and Braham 1949, Table 6).

V. Acknowledgments

VI. References

Austin, J. M. (1951). —" Compendium of Meteorology." p. 694. (Amer. Met. Soc. : Boston, Mass.)

Brunt, D. (1927).—*Quart. J. R. Met. Soc.* **53** : 30.

Byers, H. R., and Braham, R. R. (1949).—" The Thunderstorm." (U.S. Weather Bureau : Washington, D.C.)

Normand, C. W. B. (1946).—*Quart. J. R. Met. Soc.* **72** : 145.

Petterssen, S., Knighting, E., James, R. W., and Herlofson, N. (1946).—*Geofys. Publ.* **16** (10) : 1.

Priestley, C. H. B., and Swinbank, W. C. (1947).—*Proc. Roy. Soc.* A **189** : 543.

Rayleigh, Lord (1916).—*Phil. Mag.* **32** : 529.

Richardson, L. F. (1926).—*Proc. Roy. Soc.* A **110** : 709.

Scorer, R. S., and Ludlam, F. H. (1953).—*Quart. J. R. Met. Soc.* **79** : 94.

Smith, E. J. (1951).—*Quart. J. R. Met. Soc.* **77** : 33.

Sutton, O. G. (1950).—*Proc. Roy. Soc.* A **204** : 297.

von Weizsäcker, C. F. (1948).—*Z. Phys.* **124** : 614.

Workman, E. J., and Reynolds, S. E. (1949).—*Bull. Amer. Met. Soc.* **30** : 359.

Appendix I
The Form Factor c

Equations (**3**) involve a numerical factor c which is characteristic of the form of the distribution of w or T within the element. In the absence of information about this distribution, the likely magnitude of c can only be assessed by

D

assuming reasonable analytic forms for the distribution and evaluating.   The
quantity considered is supposed to be symmetrical about a maximum central
value, and using r for radial distance we shall evaluate two functions, the normal
form

(i) $$w = w_0 e^{-r^2/R^2},$$

which has a finite value (w_0/e) and finite slope at the limits ($r = R$), and

(ii) $$w = w_0\left(1 - \frac{r^2}{R^2}\right),$$

which falls sharply to zero at the limits.   Since there is interest also in the
condition of cylindrical symmetry (columnar convection) we include under (ii)
the case where r and R represent distances from a central axis.   The results of
averaging yield :

	(i) Spherical	(ii) Spherical	(iii) Cylindrical
$\bar{w}$	$0\cdot57w_0$	$0\cdot4w_0$	$0\cdot5w_0$
$\overline{\nabla^2 w}$	$-\dfrac{2\cdot2w_0}{R^2}$	$-\dfrac{6w_0}{R^2}$	$-\dfrac{4w_0}{R^2}$

whence

$\overline{\nabla^2 w}$	$-\dfrac{4\bar{w}}{R^2}$	$-\dfrac{15\bar{w}}{R^2}$	$-\dfrac{8\bar{w}}{R^2}$

These values have led to the adoption of $c = 8$ in the one place in the paper
where such appeal is necessary, but a large measure of uncertainty remains at
present in this or other instances where conversion from values of k to values of
R is required.

THE STRUCTURE OF THE F REGION OF THE IONOSPHERE

By A. A. Weiss*

Summary

The effects of temperature cycles, of a decay coefficient non-uniform with height, and of vertical tidal drifts upon the structure of the F region of the ionosphere are considered. Special attention is devoted to the behaviour of the level at which the maximum electron density occurs.

Both decay by recombination and decay by attachment are examined. It is concluded that the diurnal and seasonal height variations of the F_2 region are not the result of vertical tides alone acting on an isothermal Chapman region. A qualitative explanation of these variations is obtained by postulating, in addition, a diurnal temperature cycle, provided the decay coefficient does not change rapidly with height. Under the alternative hypothesis, that a single ionization process forms both F_1 and F_2 regions by bifurcation when the decay coefficient changes rapidly with height, the height variations appear to require discontinuities in the height gradient of the decay coefficient ; tidal drifts are still necessary but a diurnal temperature cycle, if it exists, is not of major importance.

I. Introduction

In endeavouring to unravel the complexities of the F_2 region, there has been a tendency in the past to concentrate attention upon the maximum electron density, N_m, if only because the behaviour of N_m bears some resemblance to that of the simple Chapman region. If the departures of the level, $h_{F_2}^m$, at which N_m occurs, from the Chapman norm are large, the reason is that $h_{F_2}^m$ is much more sensitive to the parameters determining the structure of the region than is N_m, as is evident from the influence of vertical tidal drifts upon an ionized region whose relaxation time is large (Weiss 1953). Further examples are presented below. Physically, such behaviour is not unexpected, in view of the " cell " motions and the long relaxation times involved.

It is therefore to be expected that the study of height variations should yield interesting and significant results ; consequently several mechanisms, any or all of which may be effective in shaping the profile of an ionized region, are considered below from this standpoint. They are (1) temperature, (2) decay coefficient, (3) solar zenith angle, (4) vertical tidal drifts.

II. The Effects of Temperature Cycles

Lepechinsky (1951) has examined the rate of ion production in an isothermal atmosphere subject to a diurnal or seasonal temperature cycle, and his results are first reiterated in a form more suited to subsequent developments. Let the

* Department of Physics, University of Adelaide.

scale height H be $H=H_0F(\chi)$, where χ is the solar zenith angle, and choose a datum level

$$h_0^0=H_0 \ln [A\rho_0H_0F(\chi)],$$

where ρ_0 is the density at $h=0$. Then, denoting the maximum value of noon ion production by I_0^0 and defining

$$z=[h-h_0^0F(\chi)]/H,$$

the ion production formulae become

$$I=[I_0^0/F(\chi)] \exp (1-z-e^{-z} \sec \chi), \quad \ldots\ldots\ldots\ldots (1)$$
$$I_m=I_0^0 \cos \chi/F(\chi),$$
$$h(I_m)=F(\chi)(h_0^0+H_0 \ln \sec \chi),$$
$$z(I_m)=\ln \sec \chi.$$

As a representative example, a linear diurnal temperature cycle for the equator at the equinoxes may be postulated, that is,

day :
$$F(\chi)=1\pm\frac{12\mu}{\pi}\varphi,$$
night :
$$F(\chi)=1-12\mu,$$
$$\left.\right\} \quad \ldots\ldots\ldots\ldots (2)$$

where μ is a constant which specifies the rate of change of temperature with time and φ is the time in radians measured from local noon as zero, so that $\chi=\varphi$. The positive sign applies to the morning hours, when $\varphi<0$. With $\rho_0H=$constant, $h_0^0=4H_0$, and $\mu=0\cdot04$, so that H ranges from 53 to 70 km, the diurnal variation of $h(I_m)$ shown in Figure 1 is obtained. During most of the day $h(I_m)$ lies below the noon level (here 280 km), in marked contrast to normal Chapman behaviour.

Associated with a diurnal temperature cycle may be a cyclic expansion and contraction of the atmosphere as a whole. The resulting transport of cells of ionization and modifications of electron density may be taken into account in exactly the same manner as tidal cell displacements. If v be the vertical cell motion, measured positive upwards, the electron density contours are found by solving the continuity equation for electrons,

$$\frac{\partial N}{\partial t}=I-\alpha N^2-\frac{\partial}{\partial h}(Nv), \quad \ldots\ldots\ldots\ldots\ldots (3)$$

after the manner indicated by Weiss (1953). The simplest treatment is to assume that the structure of the atmosphere remains isothermal at any instant throughout the diurnal temperature cycle and that the pressure at the datum level remains constant. The laws of motion are then

$$v=\frac{h}{H}\frac{dH}{dt}, \quad \frac{\partial v}{\partial h}=\frac{1}{H}\frac{dH}{dt}, \quad \ldots\ldots\ldots\ldots (4)$$

and, with $F(\chi)$ given by (2),

$$h=h_0F(\chi), \quad v=\pm h\mu/F(\chi), \quad \frac{\partial v}{\partial h}=\pm\mu/F(\chi), \quad \ldots\ldots (5)$$

where h_0 is the noon level of a chosen cell of ionization. With the assumption of constant density at the datum level these equations of motion are modified by the addition to the velocity of a term independent of height, thus :

$$v = \left(1 + \frac{h}{H}\right)\frac{\mathrm{d}H}{\mathrm{d}t}, \quad \frac{\partial v}{\partial h} = \frac{1}{H}\frac{\mathrm{d}H}{\mathrm{d}t},$$

but the distinction is important only close to the datum level.

The continuity equation (3) has been integrated with the equations of motion (5) and (2), and with physical conditions representative of the F_2 region. The behaviour of h_m, the height at which the relative maximum electron density $\nu_m = N_m/N_0$ occurs, is illustrated in Figure 1. As before, $\rho_0 H = \mathrm{constant}$, $h_0^0 = 4H_0$, and $\mu = 0\cdot04$. N_0 is the noon maximum equilibrium electron density. The
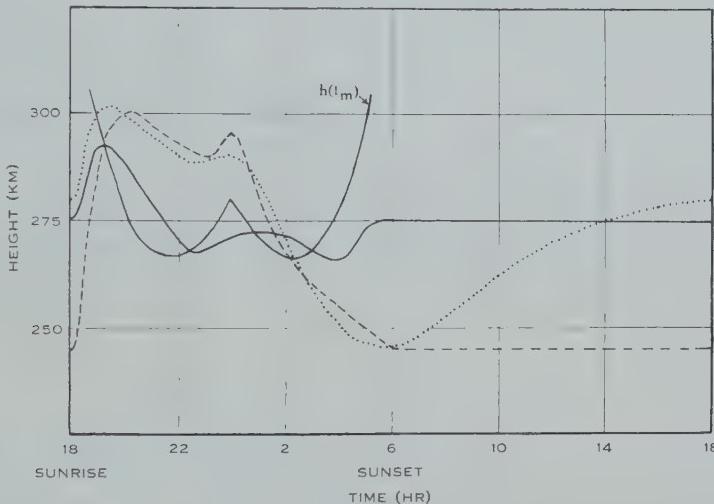
Fig. 1.—Movement of the level of maximum electron density, z_m, under diurnal temperature cycles.

$$h_0^0 = 280 \text{ km}, \quad \mu = 0\cdot04 \text{ (see text).}$$
———— No cell motion.
– – – – Cell motion, $\alpha = \mathrm{constant}$.
· · · · Cell motion, $\alpha = \alpha_0 \exp(-z/10)$.

case of no cell motion, $v = 0$, with I given by (1), is also shown. The third curve relates to a region for which the rate of decay decreases slowly with height, $\alpha = \alpha_0 \exp(-z/10)$, where α_0 is the value of the decay coefficient at the datum level $z = 0$. It is easily seen that an increase in the range of the daily temperature cycle (increase in μ) is accompanied by a rise in the noon level of h_m when cell motion is taken into account, but by a fall in noon h_m when cell motion is ignored. An increase in h_0^0/H_0 implies a rise in noon h_m relative to sunrise (and sunset) h_m.

The foregoing examples cannot be considered as representing the actual atmosphere. It is not to be expected that the upper atmosphere would in fact retain an instantaneously isothermal structure during a cycle of temperature changes. Indeed, from several independent sources (see, e.g. Gerson 1951) there is evidence of a temperature gradient between the E and the F_2 regions.

The rates of ion production in an atmosphere in adiabatic equilibrium, with temperature rising linearly with height, have been evaluated by Gledhill and Szendrei (1950); their conclusions that the larger the temperature gradient or the base temperature, the higher the level of region formation, the thicker the region, and the lower the electron density are clearly consistent with the results of Lepechinsky. Seasonal temperature changes in the upper atmosphere appear to be substantiated and a diurnal cycle may reasonably be inferred. This being the case, Figure 1 may serve to indicate the way in which a diurnal temperature cycle can influence h_m for the F_2 region. The main effect is to raise the level of h_m above that appropriate to the simple Chapman region, for which the day-time level of h_m is lower than the night-time level except immediately after sunrise. The detailed electron density profiles, not reproduced here, show that ν_m is essentially the same as on the simple Chapman theory. Further calculations show that the addition of a semi-diurnal tidal drift of electrons, whose phase is consistent with the observed night-time elevation of $h_{F_2}^m$, would suffice to account for the day-time elevation of the F_2 region, which is typical of summer conditions over most of the world.

At the level of the F_1 region the relaxation time is sufficiently short that neither tidal drifts nor thermal cell motions can materially alter electron density profiles. If a diurnal temperature cycle is present at F_1 levels, then $h_{F_1}^m$ should closely follow $h(I_m)$. The fact that the F_1 region follows Chapman behaviour appears to suggest the absence, at F_1 levels, of appreciable temperature cycles, either diurnal or seasonal.

III. The Low Attenuation Region

The F_2 region has been described by Bates (1949) as a low attenuation region, exhibiting less solar control than a normal Chapman region. According to this same author the ionizing process responsible for the F_1 region will provide a sufficient rate of ion production at higher levels to account for the F_2 region also, when regard is taken of the decrease of recombination coefficient from the F_1 to the F_2 region. In addition to the " recombination " law under which removal of electrons proceeds according to the law $dN/dt = -\alpha N^2$, attention has been directed by certain authors (e.g. Bates and Massey 1946) to the possibility that an " attachment " law, $dN/dt = -\beta N$, may apply to the F_2 region, rather than the recombination law. In respect of the attachment law an " effective recombination coefficient " may be defined by writing $\alpha = \beta/N$. Now, despite the large scatter in the measured value of the recombination coefficient (see, e.g. Bates and Massey 1946), there can be no doubt that it decreases by one and probably two or more orders across the F region as a whole, and, whether the F region originates in one ionizing process or in two, it is pertinent to examine the structure of an ionized region for which the decay coefficient is height dependent.

The characteristic of a low attenuation region is that the rate of ion production depends primarily on the concentration of the active constituent, and only secondarily on the solar zenith angle χ. Since in the Chapman ion production formula, $I = I_0 \exp(1 - z - e^{-z} \sec \chi)$, the factor $\exp(-e^{-z} \sec \chi)$ becomes

negligible for $z > 2$, the upper levels of the isothermal Chapman region will serve as an example of a low attenuation region. The absence of tidal and thermal displacements is assumed.

(a) Recombination Law

If the recombination coefficient falls off exponentially with height, the structure of the region at any given level is determined by the equations (cf. Chapman 1931)

$$\text{day:} \qquad \frac{d\nu}{d\varphi} = q^2 - p^2\nu^2, \left.\begin{array}{c} \\ \\ \\ \end{array}\right\} \quad \cdots\cdots\cdots\cdots\cdots \quad (6)$$

$$\text{night:} \qquad \frac{d\nu}{d\varphi} = -p^2\nu^2,$$

which are the forms assumed by the continuity equation (3) in the absence of vertical drift ($v = 0$) when time is measured in terms of φ and the electron density by the relative electron density $\nu = N/N_0$. N is the electron density at any level at a given time and $N_0 = (I_0/\alpha_0)^{\frac{1}{2}}$ as before is the noon maximum equilibrium electron density; α_0 is the recombination coefficient at the level $z = 0$. $q^2 = a \exp (1 - z)$ and $p^2 = a \exp (-\gamma z)$, where γ is a constant specifying the rate of decrease of the decay coefficient with height. The adjustable parameter a is defined by $a = (1 \cdot 37 \times 10^4 N_0 \alpha_0)^{-1}$. A range of values of a from $40/\pi$ to $1/\pi$ should cover all possibilities for noon conditions for either F_1 or F_2 regions. The day-time solution of (6) is

$$\nu = \frac{q}{p}(ke^{2qp\varphi} - 1)/(ke^{2qp\varphi} + 1),$$

where k is a constant of integration. Reckoning time from sunrise as zero, then for the equinoxes the relative electron densities at sunrise and sunset are respectively

$$\nu_0 = \frac{q}{p}(k - 1)/(k + 1),$$

$$\nu_{12} = \frac{q}{p}(ke^{2qp\pi} - 1)/(ke^{2qp\pi} + 1). \quad \cdots\cdots\cdots\cdots \quad (7)$$

The night-time solution of (6) is

$$\frac{1}{\nu} = \frac{1}{\nu_{12}} + p^2\varphi,$$

and at sunrise

$$\frac{1}{\nu_0} = \frac{1}{\nu_{12}} + p^2\pi. \quad \cdots\cdots\cdots\cdots\cdots \quad (8)$$

If $qp\pi \ll 1$, it is found, by equating (7) and (8), that $k = 3 + 2\sqrt{2}$, independent of p and q. For larger values of $qp\pi$, the expression for k is more complex; these solutions are not investigated here. The electron density profiles are, for $qp\pi \ll 1$,

$$\text{sunrise:} \qquad \nu_0 = \frac{1}{\sqrt{2}} \exp \tfrac{1}{2}(1 - z + \gamma z), \left.\begin{array}{c} \\ \\ \\ \end{array}\right\} \quad \cdots\cdots\cdots \quad (9)$$

$$\text{sunset:} \qquad \frac{1}{\nu_{12}} = \frac{1}{\nu_0} - a \exp (-\gamma z).\pi.$$

(b) Attachment Law

If decay proceeds according to an attachment law, with the rate of decay falling off exponentially with height, the relevant equations determining the structure of the region are

day :
$$\frac{d\nu}{d\varphi} = q^2 - p^2\nu,$$

night :
$$\frac{d\nu}{d\varphi} = -p^2\nu.$$

q and p have the same significance as in equation (6) if the attachment coefficient at the datum level, β_0, has the value $\beta_0 = \alpha_0 N_0$. The solutions of these equations, found in a manner similar to the above, are

sunrise : $\nu_0 = \exp(1 - z + \gamma z)/[1 + \exp(p^2\pi)]$,
sunset : $\nu_{12} = \nu_0 \exp(p^2\pi)$. $\left.\right\}$ ...... (10)

These latter profiles are valid for all values of p and are limited in applicability only by the extent to which neglect of the factor $\exp(-e^{-z} \sec \chi)$ is warranted. The attachment profiles (10) for $\gamma = 0 \cdot 75$ are drawn in Figure 2.
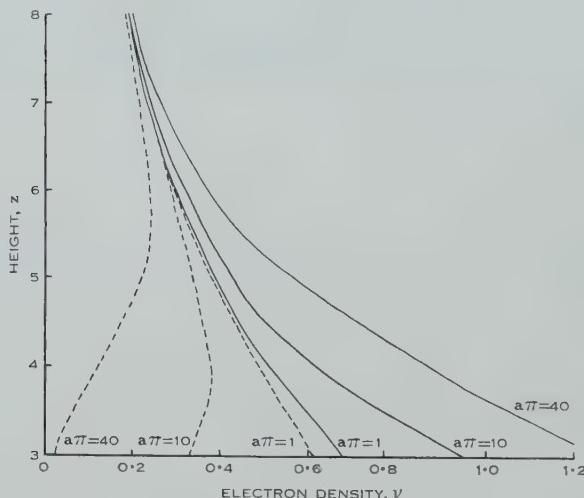
Fig. 2.—Electron density profiles for low attenuation region, $\gamma = 0 \cdot 75$.
– – – – Sunrise. ———— Sunset.

For large values of z, under either decay law, the region profiles become independent of a and hence of the absolute value of the decay coefficient. For smaller z, the attachment profiles become sensitive to a, especially for small γ, a dependence which is expected to apply also to the recombination profiles. At very high levels, the diurnal cycle in electron density tends to vanish, ν becoming independent of φ and depending only upon z, under which conditions seasonal influences can enter only through the period of illumination of the region.

It will be observed that at not too low levels or too small γ, $\nu_{\text{attach}} \rightarrow (\nu_{\text{recomb}})^2$ at sunrise For $a\pi = 1$, this is true for all γ, as can be shown by extrapolation

of (9) to $\gamma=0$ using the profiles of Chapman (1931), and also for all hours because of the small cycle in ν which is associated with small values of a. A similar relation of course holds for the equilibrium region with $d\nu/d\varphi=0$, but the conditions envisaged here are far from equilibrium.

On the hypothesis that the rate of removal of electrons falls off exponentially with height, layer formation is not possible unless $\gamma<1$, since if $\gamma>1$, $\nu\to\infty$ as $z\to\infty$. This contradicts the conclusion reached by Mitra (1952) that there may be a well-defined maximum of electron density even if $\gamma>1$. Other laws may of course be postulated which lead to layer formation at all times, e.g. $p^2=ac^n(z+c)^{-n}$, where c and n are arbitrary positive constants. Such a law will yield a maximum of electron density for all values of n.

IV. NIGHT-TIME BEHAVIOUR WITH HEIGHT-DEPENDENT DECAY

The profiles of Figure 2, for $a\pi=10$, 40 focus attention upon the fact that during the night, in the absence of ion production and other complicating factors, the level of maximum electron density of an ionized region across which the rate of decay decreases with increase in height, will invariably rise. One example, where tidal drifts have been included, has been given by Weiss (1953). The factors which determine the extent of the rise in z_m may readily be formulated.

From the night-time solution (8) for the recombination law the condition for the maximum of ν_m is easily found to be

$$\frac{1}{\nu_{12}}\frac{\partial\nu_{12}}{\partial z}=\varphi\nu_{12}\frac{\partial p^2}{\partial z}. \quad\ldots\ldots\ldots\ldots\ldots\ (11)$$

For p^2, take as before $p^2=a\exp(-\gamma z)$, and for ν_{12}

$$\nu_{12}=\exp K(1-z-e^{-z}), \quad\ldots\ldots\ldots\ldots\ (12)$$

which for $z>0$ and with K a positive constant will serve to represent the upper portions of profiles of the type drawn in Figure 2. The failure of (12) to represent the behaviour of the region for $z<0$ is of no consequence as $z_m>0$ always. Then (11) becomes

$$K(1-e^{-z_m})=\exp K(1-z_m-e^{-z_m})\varphi\gamma ae^{-\gamma z_m}. \quad\ldots\ldots\ldots\ (13)$$

For the attachment law we have, more simply,

$$K(1-e^{-z_m})=\varphi\gamma ae^{-\gamma z_m}. \quad\ldots\ldots\ldots\ldots\ldots\ldots\ (14)$$

The graphical solutions of (13) and (14), for $\varphi=\pi$, are presented in Figure 3, which shows how the extent of the rise in z_m after 12 hr depends on γ, K, and a. It will be seen that the rises of z_m are by no means small, and that the attachment law invariably gives larger rises than the recombination law. The association of small rises in z_m with large values of γ is physically self-evident. The curves have not been drawn beyond $\gamma=1$ because, as already indicated, region formation is impossible for $\gamma\geqslant1$. Some uncertainty attends these results because (12) may not adequately represent the initial conditions at sunset (although experience with complete integrations suggests that the sunset profiles above z_m are of this shape), but the conclusion appears inescapable that a small rise in z_m at night implies either a long relaxation time or a very small gradient of the decay coefficient with height.

V. THE COMPLETE REGION PROFILES

The conclusions of the preceding sections are corroborated by several integrations over 24-hr periods of the continuity equation (3) for electrons, which have been performed for the equator at the equinoxes. The results—diagrams of z_m and of ν_m—are collected in Figure 4. For the recombination law ion production follows $[1/F(\chi)] \exp(1-z-e^{-z} \sec \varphi)$ and decay the law $[1/F(\chi)]^2 e^{-\gamma z} \nu^2$ with $\gamma = 0, 0 \cdot 25$. For the attachment law, ion production is as above, and decay follows $[1/F(\chi)]^2 e^{-\gamma z} \nu$, $\gamma = 0, 0 \cdot 1, 0 \cdot 25, 0 \cdot 5$, and $0 \cdot 8$. In both cases $F(\chi)$ is given by equation (2) with $\mu = 0 \cdot 04$, and for the third term
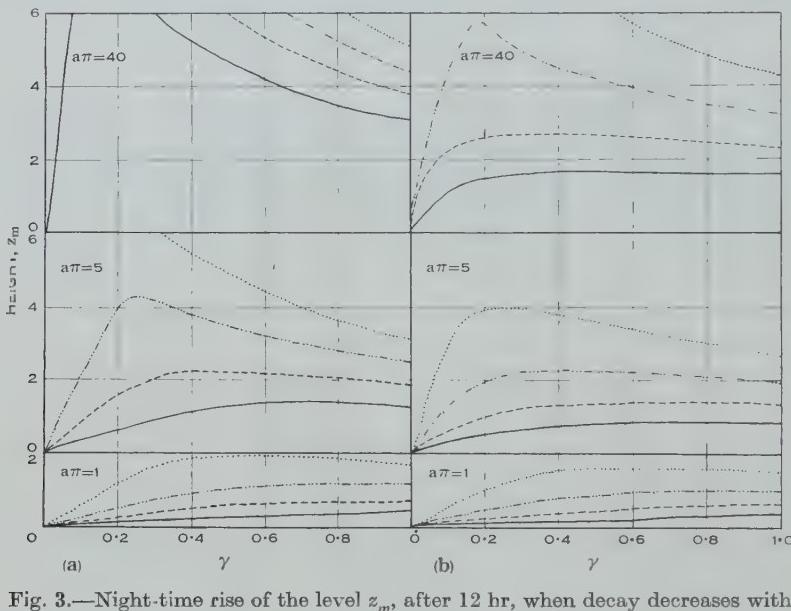
Fig. 3.—Night-time rise of the level z_m, after 12 hr, when decay decreases with increasing height. Ascending portions of some curves omitted. (a) Attachment, (b) recombination.

—————— $K = 2$.   —··—··—··— $K = \frac{1}{2}$.
– – – – $K = 1$.   · · · · · · · $K = \frac{1}{4}$.

of the continuity equation the value of $\partial v/\partial h$ given by (5) has been adopted, that is, $\partial v/\partial h = \pm \mu/F(\chi)$. The inclusion of the factors in $F(\chi)$ in these expressions admits the possibility of including thermal motions as in Section II but, as $F(\chi) \sim 1$ except near sunrise and sunset and the velocity gradient is small at all times, Figure 4 will depict with sufficient accuracy the behaviour of z_m and ν_m for the static isothermal region. Profiles for the sequence $0 \leqslant \gamma \leqslant 0 \cdot 8$ for the attachment law are shown in Figure 5.

From these diagrams the following conclusions may be drawn :

(a) A rough approximation to the profile, and even to the behaviour of ν_m, can be obtained by putting $dv/d\varphi = 0$.

(b) The movement of z_m, which is similar in form with either law of decay, cannot be inferred by putting $dv/d\varphi = 0$.

(c) Although the separation between z_m and $z(I_m)$ increases as γ increases, during the middle part of the day the level of ν_m is to be found not far from the level of the maximum rate of ion production. Such a conclusion is of course only valid for values of γ somewhat less than 1.
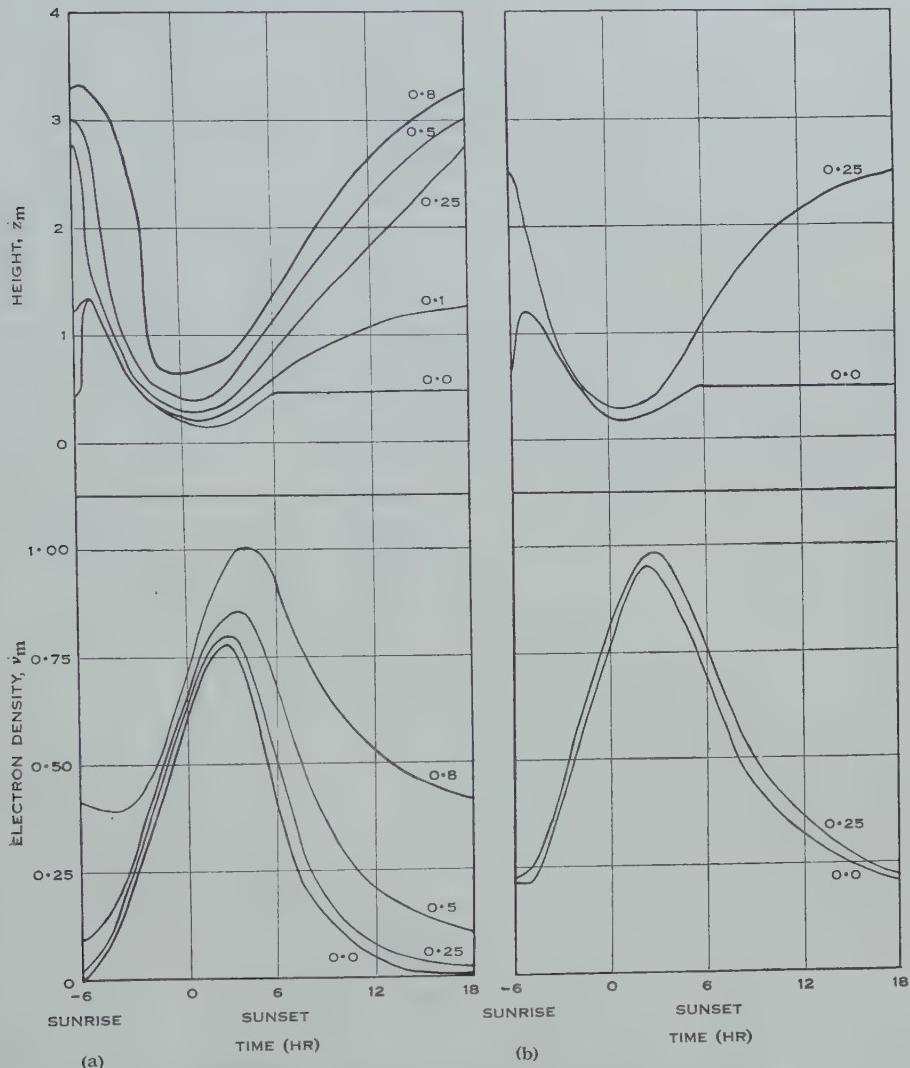
Fig. 4.—Region parameters, z_m and ν_m, when decay decreases with height. The numbers on the curves are the values of γ. (a) Attachment, (b) recombination.

(d) For the intermediate value of a chosen for the integrations $\nu_{attach} \sim (\nu_{recomb})^2$. During the middle part of the day $d\nu/d\varphi \to 0$ quite closely at low levels, whilst at higher levels the equilibrium condition is simulated (Section III). Further confirmation of this relation lies in Figure 3, with $a\pi = 5$, for which $K_{attach} = 2K_{recomb}$ gives approximately equal night-time rises of z_m.

(e) Sunrise values of ν_{attach} are much less than the corresponding values of ν_{recomb}, smaller even than implied by (d) above. If N_s is the sunset value of N_m, then with decay by recombination $N/N_s = (1 + N_s \alpha t)^{-1}$, whilst for decay by attachment, with $\beta = N_s \alpha$, $N/N_s = \exp(-N_s \alpha t)$. These two laws give equal rates of decay only if $N_s \alpha t < \frac{1}{2}$, or, for a 12-hr period, $N_s \alpha < 10^{-5}$ if N_s is measured in cm^{-3}, α in cm^3 sec^{-1}. Now, for the conditions under which these integrations were performed, $N_s \alpha \sim 4 \times 10^{-5}$ and the more rapid decay following the attachment law is not unexpected. Values of α approaching 10^{-11} are necessary before the two laws lead to equal rates of decay over a period as long as 12 hr, and under such conditions N/N_s remains quite large, $N/N_s > \frac{1}{2}$.
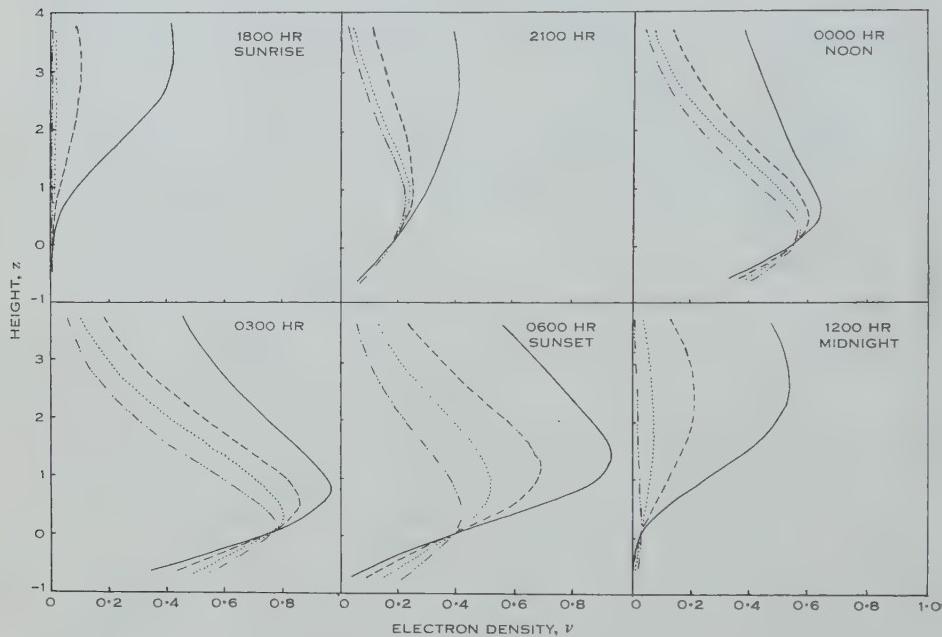
Fig. 5.—Electron density profiles when decay proceeds by attachment.

————— $\gamma = 0 \cdot 8$.   · · · · · $\gamma = 0 \cdot 25$.
– – – – $\gamma = 0 \cdot 5$.   —··—··— $\gamma = 0 \cdot 0$.

VI. DISCUSSION OF HEIGHT VARIATIONS

Height data for the F_2 region have been summarized by Appleton (1950). The idealized annual variation of height as a function of latitude relates to noon h'_{F_2} and, if a vertical tidal drift with height gradient contributes to these variations, noon $h^m_{F_2}$ may not vary with season and latitude in exactly the same manner as noon h_{F_2}. However, in examining how far these annual variations in height can be explained by vertical tides acting on a simple (isothermal) Chapman region, we may assume that the two heights are directly comparable.

According to Martyn (1948) and Mitra (1952), the tidal drift at any station may be represented as the sum of two terms, one annual and the other seasonal. Since Weiss (1953) has shown that even a considerable height gradient in the

vertical tidal drift velocity has little influence on z_m, and in any case the available data indicate that tidal drift velocities at the F_2 level are not large ($\sim$10 km/hr) it is sufficient to consider a semi-diurnal tide uniform with height, which is directly additive. If v_a and v_s are the amplitudes of the annual and seasonal components of the total vertical tidal drift and λ_a and λ_s the respective phase angles, the total tidal drift velocity is

$$v = v_a(\theta) \sin (2\varphi + \lambda_a) + v_s(\delta, \theta) \sin (2\varphi + \lambda_s + \Omega t), \quad \dots \ (15)$$

where $\theta =$ co-latitude, $\delta =$ solar declination, and $\Omega t = 30°$/month.

The seasonal variation of noon z_m is then found, after integration of (15) and incorporating the control of region height by zenith angle (second term), to be

$$w_{\text{noon}} = -\tfrac{1}{2}[v_a(\theta) \cos \lambda_a + v_s (\delta, \theta) \cos (\lambda_s + \Omega t)] + \ln \operatorname{cosec} (\theta + \delta).$$

The terms in δ will introduce a semi-annual variation at latitudes $< 23\tfrac{1}{2}°$ and an annual variation at higher latitudes, whilst the term in Ωt produces an annual variation at all latitudes. Close to the equator the seasonal tidal term $v_s(\delta, \theta)$ should be small and the zenith angle term is in phase quadrature with the actual variation of noon height. At high latitudes one would expect the zenith angle term to dominate, and this is in phase opposition with the observed variations.

The conclusion that the seasonal height variations of the F_2 region cannot be satisfactorily explained by vertical tides acting on an isothermal Chapman region is strengthened by consideration of the diurnal cycle in $h^m_{F_2}$ for a station such as Washington. Here the seasonal variation is well developed, and is related to the radical change in $h^m_{F_2}$ from Chapman-like behaviour in winter months to high day-time values in summer. It does not admit of ready explanation in terms of a seasonal cycle in the amplitude, or phase, or both, of a tidal drift. It is pertinent to mention here that Martyn (1948) has shown that the height variations of the F_1 region are consistent with small tidal displacements, which at this level do not necessarily imply small tidal drift velocities, of an isothermal Chapman region.

Apart from influencing tidal drifts and displacements of the level of the maximum rate of ion production, the seasonal cycle in the solar zenith angle may alter region heights through an annual temperature cycle. Increased summer temperatures at the level of the F region are substantiated in several different ways. It has already been shown that a combination of diurnal temperature cycle and tidal drift, with a decay coefficient independent of height, appears adequate to account for the diurnal $h^m_{F_2}$ variations in summer months ; from Section IV it is clear that this explanation only holds provided the decay coefficient remains essentially independent of height. Further, according to Lepechinsky (1951), two regions, produced in two distinct ionizing processes, have smaller noon separation in winter than in summer owing to the smaller diurnal temperature cycle in winter. However, so long as temperature cycles alone are invoked, difficulties which cannot be resolved by appeals to tidal theory may still arise in connection with the bifurcation of the summer night F region into separate F_1 and F_2 regions just after sunrise and with the failure of

evidence of the temperature changes to appear in the diurnal and seasonal cycles of $h_{F_1}^m$.

The origin of the F_1 and F_2 regions in a single ionizing process has been considered qualitatively by Martyn (1948) and Bates (1949); the former has also stressed the importance of tidal phenomena to the structure of the composite region. Mitra (1952) has drawn profiles for a region with $\gamma = 1 \cdot 5$, but as already mentioned this result has not been confirmed by the present investigation. The behaviour of z_m for $\gamma < 1$ (Fig. 4) bears no resemblance to the diurnal cycles found for $h_{F_2}^m$, nor does it admit of sufficient separation between z_m and the level of maximum ion production for bifurcation on the scale found in summer months. The night-time profiles for $\gamma \geqslant \frac{1}{2}$ (Fig. 5) can scarcely be described as parabolic, but Ratcliffe (1951) has found for the three stations he analysed that the profiles, if not parabolic at sunset, rapidly become so and retain this shape during the night. These observed profiles of course apply only to the underside of the region, below $h_{F_2}^m$.

Although region formation is not directly possible for $\gamma \geqslant 1$, two maxima of electron density may arise in a single ionizing process if there is a double discontinuity in the height gradient of the decay coefficient, with $\gamma \sim 0$ below a lower level (giving rise to an F_1 region essentially similar to the simple Chapman region) and above an upper level (giving rise to the upper portion of the F_2 region, above $h_{F_2}^m$), with $\gamma > 1$ between these two levels. It is readily seen that for large values of a (Section III), values of γ not very much greater than unity will suffice to produce the observed ratio of $N_{F_2}^m$ to $N_{F_1}^m$ of from 2 to 4, whether decay proceeds by recombination or by attachment. During the summer night-time the upper discontinuity in the decay coefficient will act as an effective barrier preventing any rise of $h_{F_2}^m$ above the sunset level, other than that due to tidal drift. During winter in high latitudes the contraction of the atmosphere as a whole will lower both discontinuities in the decay coefficient, and this, coupled with the relative elevation of the region consequent upon reduced solar zenith angle, may well result in the complete absorption of the effective solar radiation above the level of the upper discontinuity, where decay is not strongly dependent upon height, with a normal Chapman region and failure of bifurcation to appear. Some additional mechanism, such as tides, would probably be necessary to account for the major seasonal alterations in the distribution of electrons within the F_2 region, as reported by Ratcliffe (1951).

This explanation is consistent with the results of Smith (1952), who from studies of the refraction of radio stars at Cambridge has found that the integrated electron density above $h_{F_2}^m$ is normal for a Chapman region in winter and whilst somewhat above the Chapman norm during summer is certainly not excessively so.

VIII. References

Appleton, E. V. (1950).—*J. Atmos. Terr. Phys.* **1** : 106.

Bates, D. R. (1949).—*Proc. Roy. Soc.* A **196** : 562.

Bates, D. R., and Massey, H. S. W. (1946).—*Proc. Roy. Soc.* A **187** : 261.

Chapman, S. (1931).—*Proc. Phys. Soc. Lond.* **43** : 26.

Gerson, N. C. (1951).—*Rep. Progr. Phys.* **14** : 316.

Gledhill, J. A., and Szendrei, M. E. (1950).—*Proc. Phys. Soc. Lond.* B **63** : 427.

Lepechinsky, D. (1951).—*J. Atmos. Terr. Phys.* **1** : 278.

Martyn, D. F. (1948).—*Proc. Roy. Soc.* A **194** : 445.

Mitra, A. P. (1952).—*Indian J. Phys.* **26** : 79.

Ratcliffe, J. A. (1951).—*J. Geophys. Res.* **56** : 487.

Smith, F. G. (1952).—*J. Atmos. Terr. Phys.* **2** : 350.

Weiss, A. A. (1953).—*J. Atmos. Terr. Phys.* **3** : 30.

MAXWELL–WAGNER LOSS AND ABSORPTION CURRENTS IN DIELECTRICS

By B. V. Hamon*

Summary

The dielectric loss factors of a number of heterogeneous mixtures have been measured at frequencies up to 3 Mc/s. The mixtures consisted of a non-polar base material and small percentages of a slightly conducting solid or liquid " impurity ".

Two types of dispersion of the solid impurity were studied, one using discrete spheres and the other using a fine powder covering the surfaces of large grains of the base material. For the discrete spheres, good agreement with Wagner's theory was obtained. The results for the powdered solid impurity and the liquid impurity show that such mixtures can give rise to dielectric loss factors that increase slowly with decreasing frequency over a wide frequency range, and finally merge with the loss due to D.C. conductivity. The mixtures also gave rise to anomalous charging currents when direct current was applied. These currents were measured, and their relation to the A.C. loss factors indicated.

The behaviour of the mixtures under different D.C. and A.C. potential gradients was also studied. It was found that the steady-state D.C. conductivity increased with increasing potential gradient, but that the A.C. loss factor at a particular frequency did not change. An explanation is suggested for this behaviour.

The results are compared with the published data for practical dielectrics, and support the view that, at least for non-polar dielectrics, the small loss factors and D.C. charging currents frequently observed may be due to conducting impurities present as a separate phase.

I. INTRODUCTION

Several materials of practical importance have loss factors that are small at radio and audio frequencies but increase considerably with decreasing frequency below this range. The mechanism responsible for this loss is not clearly understood. Since loss of this type is found in materials of very different natures, it is probable that a number of mechanisms are involved. In some cases (e.g. glass, and perhaps some polar materials) it is likely that the loss is an inherent property of the material; but in other cases, and especially with non-polar materials, it is usually thought that the loss is due to impurities. However, very little is known about the nature, quantity, or distribution of impurities that would lead to dielectric loss of the observed type, and few investigations have been made on systems of known composition and structure.

The present paper gives the results of an experimental investigation of certain non-polar materials to which have been added small quantities of slightly conducting " impurities ". In all cases the added impurity formed a separate

---

* Division of Electrotechnology, C.S.I.R.O., University Grounds, Sydney.

phase, so that the systems studied were heterogeneous mixtures. Dielectric loss due to slightly conducting impurities present as a separate phase in a material is now referred to as " Maxwell-Wagner loss ", since it was first investigated theoretically by Maxwell (1892) and Wagner (1914). A brief summary of the theoretical properties of Maxwell-Wagner systems will simplify the discussion of the experimental results.

For the simple case of spherical particles (conductivity σ_2 ohm^{-1} cm^{-1}, relative permittivity ε_2') distributed at random in a base material of relative permittivity ε_1', Wagner (1914) showed that the loss factor* of the mixture is given by

$$\varepsilon'' = \frac{9q(\varepsilon_1')^2}{2\varepsilon_1' + \varepsilon_2'} \cdot \frac{f/f_{max}}{1 + (f/f_{max})^2}, \quad \cdots\cdots\cdots\cdots (1)$$

where

$$f_{max} = 1 \cdot 8 \times 10^{12} \ \sigma_2/(2\varepsilon_1' + \varepsilon_2'), \quad \cdots\cdots\cdots\cdots (2)$$

q is the volume fraction of conducting material present in the mixture, and f is the frequency in c/s. The equations hold only for values of q so small that interaction between the fields of neighbouring particles can be neglected. Expressions similar to (1) and (2) were obtained by Sillars (1937) for the case of spheroidal particles whose unique axes are parallel to the applied field.

Figure 1 shows the variation of ε'' with frequency for spheres and prolate spheroids with two different axis ratios. The curves were calculated from the equations of Wagner and Sillars, using the following constants :

$$\varepsilon_1' = 2\cdot3, \ \ \varepsilon_2' = 5\cdot0, \ \ \sigma_2 = 10^{-6} \ \text{ohm}^{-1} \ \text{cm}^{-1}, \ \ q = 0\cdot01.$$

The dashed line in Figure 1 shows the loss factor that would be obtained if the same impurity were in the form of one or more uniform cylinders extending from one electrode to the other. Provided $\varepsilon_2' \geqslant \varepsilon_1'$, as would usually be the case, this line represents an upper limit to the loss factor at any frequency, and for any shape and distribution of the impurity. The equation for the dashed line is

$$\varepsilon'' = 1\cdot8 \times 10^{12} \cdot q\sigma_2/f, \quad \cdots\cdots\cdots\cdots (3)$$

or more generally,

$$\varepsilon'' = q \cdot \varepsilon_2'', \quad \cdots\cdots\cdots\cdots\cdots (4)$$

where ε_2'' is the loss factor of the impurity itself.

The importance of the shape of impurity particles in a Maxwell-Wagner system can be clearly seen from Figure 1, since all curves are for the same amount of a given impurity. Uneven distribution of particles of a given shape may be equally important. When only the nature and amount of impurity are known, the only quantitative estimate of the properties of a mixture that appears

---

* ε' and ε'' are respectively the real and imaginary parts of the complexr elative permittivity. ε'' is related to the conductivity σ and the frequency f by the equation

$$\sigma = 0\cdot55 \ \varepsilon''f \times 10^{-12} \ \text{ohm}^{-1} \ \text{cm}^{-1},$$

and to the loss angle δ by the equation

$$\varepsilon'' = \varepsilon' \ \tan \ \delta.$$

possible is the loss factor at frequencies greater than f_{max} (equation (**2**)). The experimental results given later have been assessed in this way.

The materials used were chosen for experimental convenience, rather than as being typical of materials or impurities of practical importance. The non-polar base materials used were paraffin wax, n-docosane (a pure long-chain saturated hydrocarbon, $C_{22}H_{46}$) and polytetrafluoroethylene (P.T.F.E.). The latter was used in the form of the spongy powder supplied for moulding.

Results are given for one solid and one liquid impurity, although a number of other materials were used as impurities during the investigation (Hamon and Meakins 1950). The impurities for which results are reported are n-decyl
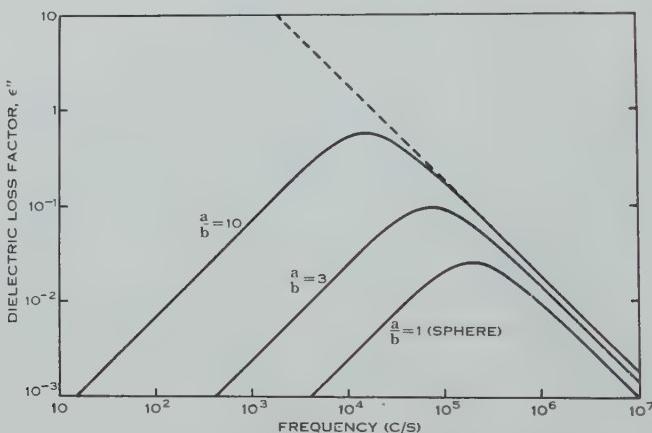
Fig. 1.—Dielectric loss factor as a function of frequency for spherical particles and for prolate spheroids with their long axes parallel to the field, calculated from the equations of Wagner and Sillars. a/b = ratio of axes.

alcohol ($C_{10}H_{21}OH$, m.p. 6·4 °C), and copper phthalocyanine. The conductivities of the samples of these materials used in the present programme were of the order of 10^{-8} and 10^{-6} ohm^{-1} cm^{-1} respectively.

The absorption and conduction currents obtained when a D.C. potential was applied to the samples were measured, as well as the more usual A.C. properties. For convenience in correlating the A.C. and D.C. results, the currents measured on direct current have been converted to loss factors at very low frequencies (Hamon 1952) and plotted on the one graph with the directly observed A.C. loss factors.

II. Materials and Procedure
(a) Materials

The n-decyl alcohol was obtained from commercial material by vacuum distillation through an efficient fractionating column. n-Docosane from a previous programme was purified by recrystallization before use. The paraffin wax and P.T.F.E. powder were used as received, but selection on the basis of electrical tests was necessary for the wax, as not all samples showed sufficiently

low dielectric loss. The copper phthalocyanine powder was a commercial product and was not purified before use.

For each of the three solids used as base material (paraffin wax, n-docosane, and P.T.F.E.) the measured loss factor was less than 0.001 at all frequencies from 0.0001 c/s to 100 kc/s.

(b) Preparation of Samples

The method of mixing depended on the nature of impurity and base, and will be described in Section III. After mixing, the material was in all cases pressed into a flat circular disk by a hydraulic press.

The compositions are expressed as percentages by volume.

(c) Electrical Measurements

The disks were measured between mercury electrodes of the type described by the Electrical Research Association (1933). On several occasions the results obtained with these electrodes were checked against those obtained with aluminium foil electrodes, which were pressed onto the sample when in the mould. No significant differences were found, either on direct or alternating current. The measuring techniques used are given in earlier publications (see, e.g. Dryden and Welsh 1951 ; Hamon 1952).

All measurements were carried out in shielded enclosures at a temperature of 20 °C. It was estimated that the loss factor values deduced from bridge measurements were not in error by more than ± 5 per cent. or ± 0.001, whichever was greater, but the values deduced from D.C. measurements might have been in error by ± 10 per cent. The measurements of D.C. conductivity at different potential gradients, reported in Section III (e), were made in a bridge circuit so that more accurate relative values could be obtained.

The conductivity of the n-decyl alcohol was measured in a conductivity cell with bright platinum electrodes. The cell was calibrated with 0.01N and 0.001N potassium chloride solutions.

Double logarithmic coordinates have been used for the graphs of loss factors as functions of frequency. With these coordinates, steady-state D.C. conduction corresponds to a straight line with a slope of $-45°$ at the low frequency end of the curve. To save space, experimental points in this region corresponding to loss factors greater than 100 have not been shown on the curves.

III. Results and Discussion
(a) Discrete, Solid Particles of Impurity

It was thought worth while, as a basis for work on more complicated mixtures, to prepare a mixture containing discrete spherical particles of impurity, and to compare the measured values of loss factor with those calculated from equations (1) and (2). Sillars (1937) prepared such a mixture by dispersing water in paraffin wax, but numerical agreement between observed and calculated results was not good.

The material used as impurity in the present work was copper phthalocyanine. The powder was pressed into a disk and its relative permittivity and

loss factor measured at different frequencies (Fig. 2). The disk, which was hard and brittle, was crushed and the resulting particles sieved to obtain a fairly uniform size. The particles were then ground to approximately spherical shape by a compressed air device (Bond 1951). After further sieving to remove dust formed by the grinding, the particles were mixed with paraffin wax, the volume concentration being 0·62 per cent.

Figure 3 shows the loss factor measured on a disk sample pressed from the above mixture, together with the values calculated from equations (1) and (2). Since neither σ_2 nor ε_2' was sufficiently independent of frequency, it was necessary
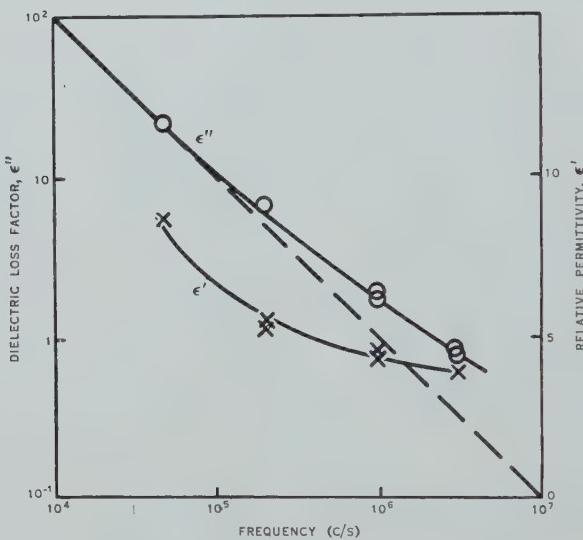
Fig. 2.—Dielectric loss factor and relative permittivity, as functions of frequency, for copper phthalocyanine.

to use different values at each frequency at which a calculated value of the loss factor of the mixture was required. The values were obtained from Figure 2, using the relation

$$\sigma = 0.55\varepsilon'' f \times 10^{-12}. \qquad \ldots\ldots\ldots\ldots\ldots \quad (5)$$

The measured value of ε_1', the relative permittivity of the wax, was 2·30. The curve calculated in this way is shown by the dashed line in Figure 3, and is slightly broader than the curves in Figure 1. The agreement between the observed and calculated values is satisfactory except at frequencies below about 5×10^4 c/s, where the observed values are appreciably greater than the calculated ones. This discrepancy may be due to slight lack of uniformity in the properties of the pressed disk, or to the fact that the particles were not exactly spherical. A photograph of part of the disk is reproduced in Plate 1, Figure 1.

Similar experiments were carried out using particles that had not been ground. The shape, height, and position of the loss factor curve were not altered appreciably by using particles of irregular shape. These tests also showed that

the loss factor is approximately linear with concentration, at least up to concentrations of about 3 per cent. This is to be expected, as it is unlikely that interaction between the electric fields of neighbouring particles will be significant at such low concentrations.

(b) Finely Powdered Impurity at the Boundaries of Large Grains

A model that should more nearly approximate to some practical materials can be made by covering large grains of the base material with an impurity in the form of a fine powder. A mixture of this kind can be expected to show a steady-state D.C. conductivity if the covering of each grain is sufficiently complete to produce conducting paths that are continuous from one electrode to the other.
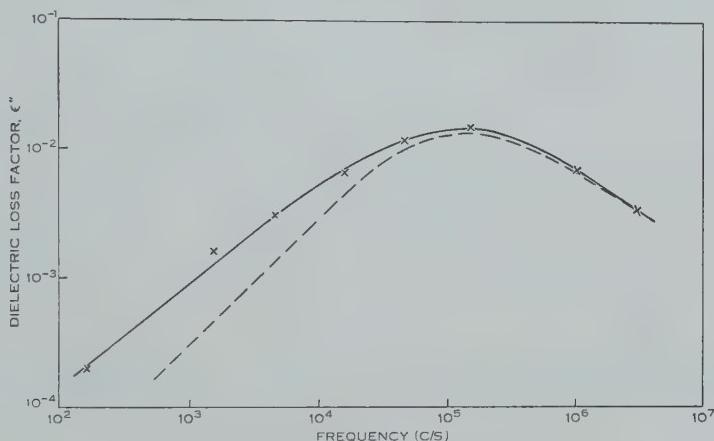
Fig. 3.—Dielectric loss factor as a function of frequency for discrete spherical particles of copper phthalocyanine embedded in paraffin wax. Solid curve, experimental; dashed curve, calculated from Wagner's equation.

For a given concentration, the D.C. conductivity should depend on the ratio between the diameters of the particles of base material and impurity, being greater for large ratios. In general, only a small fraction of the impurity will be favourably placed for D.C. conduction and the greater part of the impurity will only be able to contribute to energy loss in an A.C. field.

Figure 4 shows loss factor-frequency curves for three mixtures of copper phthalocyanine powder and paraffin wax. In preparing these samples the wax was crushed and sieved to get particles whose sizes were between about 0·05 and 0·15 cm. These particles were mixed with copper phthalocyanine powder (particle size about 10^{-4} cm) by shaking. The mixtures were pressed in the usual way. A photomicrograph of a thin section of the 0·62 per cent. sample is shown in Plate 1, Figure 2. The concentration of impurity at grain boundaries can be seen clearly.

The results shown in Figure 4 are of the general shape to be expected from the above considerations. The two higher concentrations, 0·62 and 1·9 per cent., show steady-state D.C. conductivity at frequencies below 1 c/s and 0·001 c/s

respectively, while in the 0·19 per cent. sample the concentration is too low for
D.C. conductivity.   At frequencies higher than f_{max} $(1\cdot4\times10^5$ c/s, see Fig. 3)
all three curves approach the limit curves discussed in Section I, which in this
case are respectively 1·9, 0·62, and 0·19 per cent. of the loss factor curve for
the copper phthalocyanine in bulk (Fig. 2).   Except in this high frequency
region, the loss factor is no longer linear with concentration, as was the case
with discrete particles.
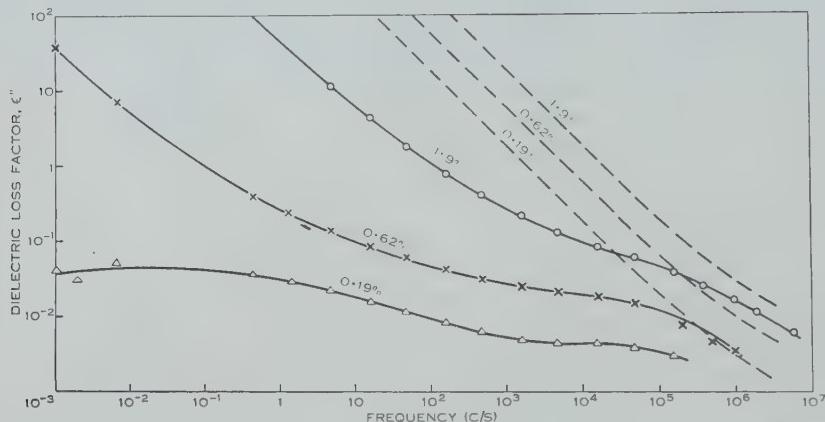
Fig. 4.—Dielectric loss factor as a function of frequency for three percentages of
copper phthalocyanine powder mixed with paraffin wax.   The dashed curves give
the theoretical upper limit to the loss factor.

(c) Liquid Impurities

Results similar to those in Figure 4 were obtained when liquid long-chain
alcohols were mixed with either paraffin wax, docosane, or P.T.F.E. powder.

Figure 5 shows the results for n-decyl alcohol mixed with P.T.F.E. powder
at three concentrations.   To ensure uniform distribution of the alcohol, it was
dissolved in pure dry light petroleum before being added to P.T.F.E.   The light
petroleum was then evaporated off until the weight was constant.

Since the alcohol is colourless, the nature of its distribution in these disks
is not as well known as in the case of the copper phthalocyanine-wax mixtures.
It appears certain, however, that the alcohol forms a film on the surfaces of the
P.T.F.E. grains, in much the same way as the copper phthalocyanine powder
covers the wax grains.   The formation of a solid solution, or appreciable penetra-
tion of the P.T.F.E. by the alcohol molecules, is not to be expected in view of the
inert nature of the P.T.F.E. towards all normal solvents.

The increasing dependence of loss factor on concentration as the frequency
decreases is clearly shown in Figure 5.   At 1 c/s, for example, the loss factor
increases by a factor of about 600 when the concentration increases by a factor
of 9, while at 10^4 c/s the corresponding increase in loss factor is only about 30
times.   From another point of view, the higher the loss factor at a particular
frequency the steeper the curve of loss factor as a function of frequency.   A

point of similarity with commercial materials is that an appreciable part of each of the curves of Figure 5 can be approximated by the empirical equation

$$\varepsilon'' = \varkappa f^{-\alpha},$$

where $\varkappa$ and α are positive constants (see, e.g. Hartshorn and Ward 1935).

For frequencies up to 10^5 c/s, the conductivity and relative permittivity of n-decyl alcohol are independent of frequency. The conductivity, however, depends on the degree of purity, and increases appreciably even when stored in a stoppered bottle in a desiccator. The conductivity of the alcohol used for the three samples whose properties are shown in Figure 5 was $9 \cdot 6 \times 10^{-9}$ ohm^{-1} cm^{-1},
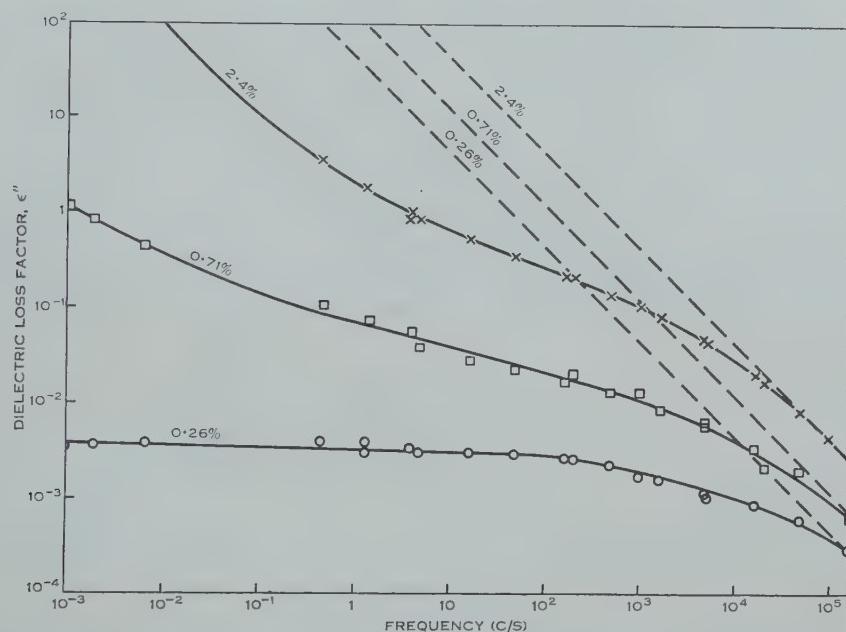
Fig. 5.—Dielectric loss factor as a function of frequency for three concentrations of n-decyl alcohol in P.T.F.E. The dashed lines give the expected upper limit to the loss factor.

and the limit curves shown in the figure are based on this value. The value of f_{max}, the frequency at which maximum loss would be expected for discrete spherical particles of decyl alcohol in P.T.F.E., is $1 \cdot 5 \times 10^3$ c/s. It is seen from Figure 5 that the loss factor falls steeply with increasing frequency above f_{max}, and that the limiting values are reached, or even slightly exceeded, at a frequency of 10^5 c/s. This point will be discussed further in relation to the results shown in Figure 6.

To compare different base materials, the results for mixtures of 1 per cent. by volume of decyl alcohol with P.T.F.E., paraffin wax, and docosane are shown in Figure 6. The P.T.F.E. curve was obtained from the results given in Figure 5 by interpolation between the $2 \cdot 4$ and $0 \cdot 71$ per cent. curves, using the formula $\varepsilon'' = A\theta^n$, where θ is the concentration and A and n are constants at a particular

frequency. The *n*-docosane and paraffin wax samples were prepared either by mixing the liquid alcohol with the molten base material, allowing the mixture to solidify, then crushing and pressing in a mould, or, more simply, by grinding the alcohol and base material with a mortar and pestle at room temperature and then pressing. The two techniques gave essentially the same result. Samples were reproducible to within about ± 20 per cent. in loss factor, as can be seen from the paraffin wax curve which has been drawn through points obtained from three different samples. Some mutual solution of alcohol and base is to be expected in the case of paraffin wax and docosane, but the fact that the loss factor curves are sensibly independent of the method of mixing, and are of the same general shape as the P.T.F.E. curves, shows that the extent of this mutual solution is probably small.
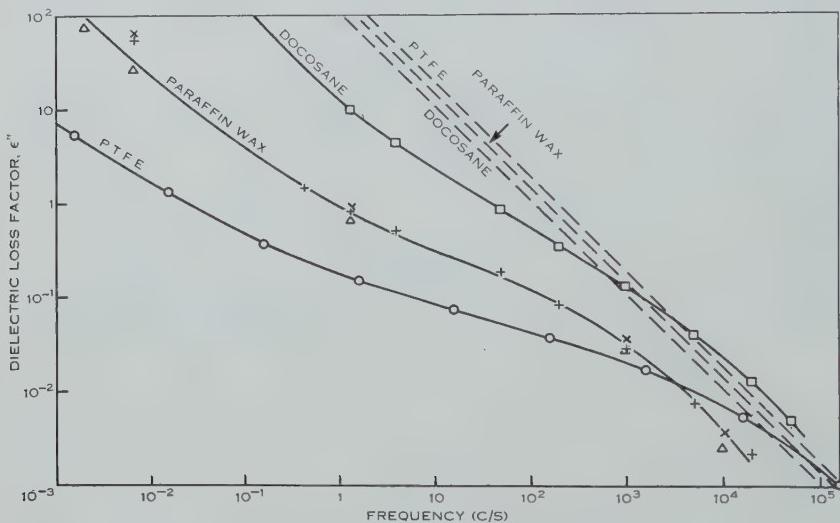
Fig. 6.—Dielectric loss factor as a function of frequency for 1 per cent. mixtures of *n*-decyl alcohol with three different base materials. The dashed curves give the theoretical upper limit to the loss factor.

Separate limit lines have been shown in Figure 6 for each sample, since the conductivity of the alcohol was slightly different in each case.

The loss in the *n*-docosane sample is appreciably higher at all frequencies than that in either the wax or P.T.F.E. samples, and is over twice as great as the expected upper limit at high frequencies. This suggests that the conductivity of the decyl alcohol in the mixture is greater than the bulk value. This could be due to orientation of the alcohol molecules as a result of adsorption on the docosane.

A similar effect is present to a smaller extent in the P.T.F.E. samples, since the difference in relative permittivities of the alcohol and P.T.F.E. (8 and $1 \cdot 9$ respectively) should prevent the loss factor becoming equal to the upper limit at any frequency.

The low value of loss factor at low frequencies in the P.T.F.E. sample is considered to be due to the large surface area of the P.T.F.E. powder. Although the grains of the powder were fairly large, each grain had a spongy texture.

It is interesting to note that loss factor curves similar to those shown in this and the preceding section have been obtained with a number of freshly ground crystalline materials (Dryden and Meakins 1953). In these systems the dielectric loss is thought to be due to conducting layers of water or aqueous solution on the freshly exposed surfaces of the crystals. Many commercial insulating materials also give dielectric absorption of this type (see, e.g. Massachusetts Institute of Technology 1948).

(d) *Anomalous Charging Currents*

The anomalous D.C. charging currents, from which the loss factors below $0 \cdot 1$ c/s were calculated, have not been considered separately in this section. In general, these currents varied with time according to the empirical law $i \propto t^{-n}$, where n is a constant for a particular sample. This empirical law has been found to hold for a large number of insulating materials.

Special mechanisms, such as the formation of space charges, and contact imperfections, have been suggested in the past to account for the anomalous currents. At least in the present cases, however, it appears simpler to assume that the anomalous currents are only another result of the heterogeneity that causes loss in A.C. fields.

(e) *The Effect of Potential Gradient*

It is well known (see, e.g. Hartshorn and Rushton 1934, 1938) that the steady-state D.C. conductivity of many commercial insulating materials increases with the applied potential gradient even at low values of the gradient, while the effective A.C. conductivity is usually independent of potential gradient, at least up to about half the breakdown value. Reversible absorption currents are usually proportional to the applied potential so that the effective " absorption conductivity " is independent of applied voltage.

On testing a number of mixture samples at different applied voltages, the same behaviour was observed. Figure 7 shows the variation of D.C. conductivity with applied potential gradient for one sample with liquid impurity (n-decyl alcohol) and one with solid impurity (copper phthalocyanine powder). In both cases the conductivity increases with the potential gradient. It was observed that the fractional increase in conductivity, for a given field change was smaller the greater the concentration of impurity. This is to be expected, since the conductivity of the impurity in bulk (n-decyl alcohol or copper phthalocyanine) does not depend on the applied potential gradient.

Figure 7 shows that the conductivity of the sample containing decyl alcohol can be represented by the empirical equation

$$\sigma = \sigma_0 \, (1 + mx),$$

where σ_0 and m are constants, and x is the potential gradient in the material. This relation is of the same form as that found for cellulose acetate and varnished cloth by Hartshorn and Rushton (1934, 1938).

Hartshorn and Rushton (1938) discussed a number of mechanisms for the dependence of D.C. conductivity on potential gradient, such as electrophoresis, movement of liquid in capillaries, and ionization by collision. None of these mechanisms seems appropriate for the mixture of paraffin wax and copper phthalocyanine (Fig. 7). A simpler mechanism is suggested in this case, namely, that the field dependence is due to the "breakdown" of small gaps in otherwise continuous paths of impurity. The nature of the breakdown process cannot be specified in detail, but the fact that the conductivity-potential gradient curve is reversible within the limits of measurement implies that no permanent physical or chemical change takes place when the gaps become conducting.
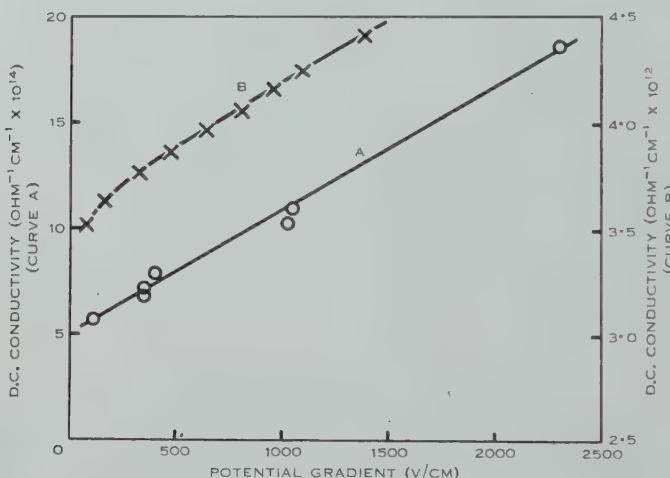
Fig. 7.—D.C. conductivity as a function of potential gradient for 1 per cent. n-decyl alcohol in paraffin wax (curve A) and for 1·9 per cent. copper phthalocyanine powder in paraffin wax (curve B).

Measurements of D.C. conductivity as a function of potential gradient were also made on a number of chemically pure organic materials. Some dependence was found, but it was generally less marked than in the mixtures discussed above. The observed field dependence in organic materials may be due to the presence of traces of moisture, or of different crystal modifications (Hamon and Meakins 1952).

IV. Conclusion

The loss factor-frequency curves in Figures 4, 5, and 6 show that a loss factor that increases slowly with decreasing frequency over a wide frequency range can be due to an impurity dispersed as a separate phase at the boundaries of grains of an otherwise loss-free material. The results obtained indicate something of the complicated relation between the properties and amount of impurity, its dispersion, and the electrical effects it produces. Although non-polar base materials were used, there is evidence that the n-decyl alcohol is in some cases adsorbed onto the base material, and that its electrical properties are altered by this process. In materials of practical importance, particularly

MAXWELL–WAGNER LOSS IN DIELECTRICS
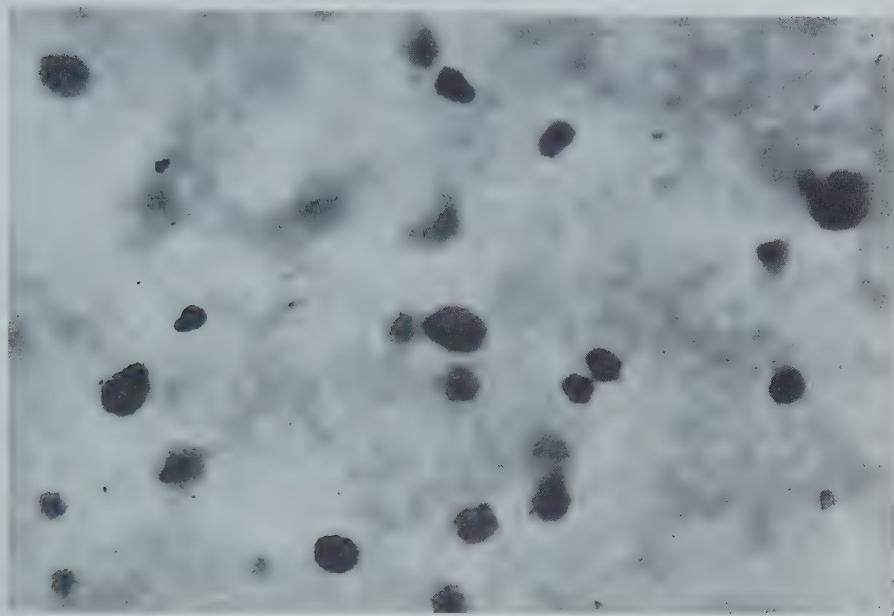
Fig. 1.   Photomicrograph of a section of a sample containing discrete spherical particles.
×15.

Fig. 2.—Photomicrograph of a section of a sample containing 0·62 per cent. copper
phthalocyanine powder in paraffin wax, showing concentration of the powder at grain
boundaries.

with polar materials exposed to atmospheric moisture, it is probable that adsorption plays an even more important part.   In such cases one can no longer use the idea of a simple mixture of components, each retaining its bulk electrical properties.

V. ACKNOWLEDGMENTS

The author wishes to acknowledge the collaboration of Dr. R. J. Meakins, Division of Electrotechnology, C.S.I.R.O., and to thank Miss M. C. Clark of the same Division for carrying out most of the measurements, and Mr. C. Healy, Amalgamated Wireless (Australasia) Ltd., for the supply of polytetrafluoro-ethylene powder.

VI. REFERENCES

BOND, W. L. (1951).—*Rev. Sci. Instr.* **22** : 344.
DRYDEN, J. S., and MEAKINS, R. J. (1953).—*Nature* **171** : 307.
DRYDEN, J. S., and WELSH, H. K. (1951).—*Aust. J. Sci. Res.* A **4** : 616.
ELECTRICAL RESEARCH ASSOCIATION (1933).—*J. Instn. Elect. Engrs.* **72** : 252.
HAMON, B. V. (1952).—*Proc. Instn. Elect. Engrs.* IV **99** : 151.
HAMON, B. V., and MEAKINS, R. J. (1950).—*Nature* **166** : 29.
HAMON, B. V., and MEAKINS, R. J. (1952).—*Aust. J. Sci. Res.* A **5** : 671.
HARTSHORN, L., and RUSHTON, E. (1934).—*J. Instn. Elect, Engrs.* **75** : 631.
HARTSHORN, L., and RUSHTON, E. (1938).—*J. Instn. Elect. Engrs.* **83** : 315.
HARTSHORN, L., and WARD, W. H. (1935).—*J. Instn. Elect. Engrs.* **77** : 723.
MASSACHUSETTS INSTITUTE OF TECHNOLOGY (1948).—" Tables of Dielectric Materials ", Vol. 3, Tech. Rept. No. 10.
MAXWELL, J. C. (1892).—" Electricity and Magnetism." 3rd Ed., p. 452. (Oxford Univ. Press.)
SILLARS, R. W. (1937).—*J. Instn. Elect. Engrs.* **80** : 378.
WAGNER, K. W. (1914).—*Arch. Elektrotechn.* **2** : 371.

PROGRAMME DESIGN FOR THE C.S.I.R.O. MARK I COMPUTER

I. COMPUTER CONVENTIONS

By T. Pearcey* and G. W. Hill*

Summary

The organization of the C.S.I.R.O. Mark I computer is described. The code system is of a two-address type in which each operation called by an " order " or " command " is thought of as a quality of a transfer of the content of one " register " to another. Each command indicates a source and a destination. A special notation of mnemonic type is adopted for written programmes as an aid in the various stages of programme design. The written programme may be translated, by aid of a special encoding key-punch, from the written sheet onto paper tape, the medium through which the machine accepts data.

I. Introduction

The basic principles of programme construction for automatic computers have been known for some time, at least since the attempt by Charles Babbage to construct an analytical engine. These principles apply to modern automatic computers, and only the details differ from machine to machine. It is one of the aims of the machine designer to make the programming procedures as simple as possible for users.

Two technical developments, the use of electronic techniques and of rapid access " variables stores ", have made fully automatic computers possible in recent years. The introduction of electronic techniques has increased the computing rate by at least a thousandfold beyond the speeds attainable by electromechanical systems using punched cards and relay equipment. This has allowed of great simplification in the organization of modern computers, the main being that sufficient speed can be maintained by performing at one time only *one* of a small class of different actions, such as addition, subtraction, reading new data, recording results, and transferring numbers and so on.

Automatic computers, particularly those operating at electronic speeds, require an extended set of single steps, known as a " programme ", which controls the course of a calculation. Thus, each step, called a " command ", must be expressed as one of the restricted class of actions which the computer is capable of performing.

Until recently automatic computers have been provided with programmes from punched cards or paper tape, each command being read from the card or tape and performed immediately, whereupon the computer proceeds to the next

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

card or position of the tape. The development of rapid access " stores " or blocks of number registers of large capacity allows the entire programme to be supplied to the computer in the form of sequences of numbers, and entered into the same store that is used for holding the incidental computing data etc. This makes it possible to subject the programme itself to changes as the computer progresses through the calculations.

The facility for a computer to change its current programme greatly reduces the number of commands required and the effort needed to construct the programme and code it into a suitable medium for transfer into the computer. Part of any programme, constructed according to this principle, must be devoted to controlling the changes to be made to itself, so that useful computing speed is somewhat reduced. The extra flexibility obtained far outweighs the disadvantages.

The problems of design and construction of modern computers have been discussed more frequently and at much greater length than have the problems of their programming and use. Only in the case of one particular computer, the EDSAC (Wheeler 1950 ; Gill 1951 ; Wilkes, Wheeler, and Gill 1951), has much been published on the latter subject.

The programming techniques developed for use in the EDSAC bear close resemblance to those developed for the present machine. This is so particularly in the use of basic programme elements or " sub-routines " which are designed for frequently repeated sets of operations, and in some of the techniques adopted for simplifying their method of use and incorporation into complete programmes.

Although the techniques used for EDSAC render programming work remarkably simple, no detailed knowledge of the machine being needed, it is believed that in some ways the method developed here is made even easier, and the demand on store space even smaller, by the incorporation into the machine of certain special features.

The terminology adopted here closely resembles that used by the Cambridge Mathematical Laboratory. Terms not well known or not immediately obvious will be defined when first introduced.

The initial aim of the programmer is of course to obtain the accuracy he needs in the results required, and this may decide for him the particular methods by which the computation is made. The programme will usually be compiled so as to use as little store space as possible, at the same time maintaining an adequate rate of output of results. These are very broad principles and subject to wide variations. Thus, a programmer designing a basic component or sub-routine will tend to occupy the least possible store space, almost irrespective of its speed, whilst a programmer designing a programme for a full-scale calculation using sub-routines is likely first to ensure that his demands do not exceed the store capacity and then choose the method requiring least machine time.

Detailed programming techniques adopted for any particular machine depend largely upon the " address system ", that is, the manner in which the machine interprets a command as a meaningful operation. Any programme implies a sequence of such operations, and must be written down in a manner easily read and understood. A programme initially designed on paper must

F

be transcribed onto some medium which is directly accepted by the machine, usually punched paper tape or punched cards, although magnetic tapes and wires are in use by certain machines. In our case the machine accepts paper tape although punched cards are used as an auxiliary medium. It is essential for simplicity in the last stage of programming, the stage which may be called " coding ", that the written programme be coded directly into the medium without further writing or extra mental effort.

In the present paper the organization of the machine is described only so far as is necessary for the understanding of the programming method, and this is followed by a description of the address system. The basic actions of the machine are listed in detail in Tables 2 and 3, together with a notation used in the construction of written programmes, which, although not ideal, appears to be suitable. The manner of coding from the written sheet is described, but a description of details of the construction of complete programmes will be given in Part II (Pearcey and Hill 1953).

For recording written programmes it is convenient that the notation symbols be available on a typewriter.

II. The Structure of the Computer

The C.S.I.R.O. Mark I computer is of the electronic serial-transfer type, and operates entirely in the binary scale. The programmer need know very little about the details of the machine but a general picture of its structure and its mode of operation is useful.

The stores, those components which hold the programme of commands and the data operated upon, are of two kinds. First, a main or high speed store (access time 1 msec), capable of storing 1024 separate data or " words " or basic groups of digits, is used always to hold the programme currently in use and incidental data as desired. This store consists of a group of mercury-filled acoustic delay lines each holding 16 words placed end to end. The register positions or " locations " in the store are numbered serially from 0000 to 1023.

The second store, of slow access (10 msec), is divided into four groups each of a capacity of 1024 separate numbers and is of the magnetic drum type. The high speed store is of the " volatile " type, that is, the numerical content is destroyed whenever the power is removed ; in the magnetic drum type this is not so and information is retained.

The arithmetical unit consists of a group of acoustic delay registers known as registers A, B, C, H, and D. The main arithmetical register is A, and has a capacity of one word only, as also have B and C. Register H, however, stores only half the digit capacity of a full number, whilst D can store 16 words end to end.

Other registers, not part of the arithmetical equipment, are the " input register " from which data fed to the computer may be read into the stores or the arithmetical components ; the " output register " via which results may be recorded ; and " constant registers ", two of which have full digit capacity for a complete word and may be adjusted manually by sets of switches. Three others provide single unit digits in prescribed places.

The operation of the computer is controlled according to the numerical content of three registers. The " sequence register " specifies from which location in the high speed store a command is to be withdrawn for use. The " interpreter register " receives and holds the command in use. The " store control register " indicates which location in one or other of the stores is prepared for immediate use.
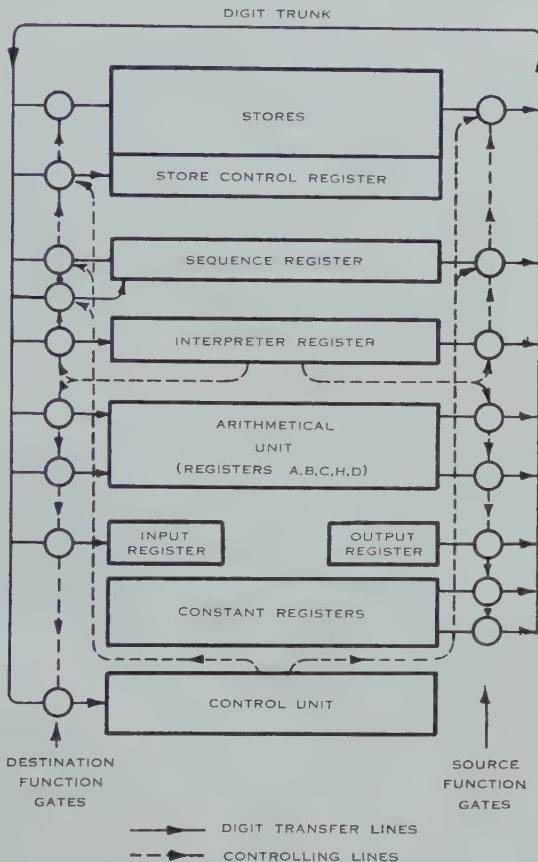
Fig. 1.—Block schematic diagram of the computer.

Figure 1 shows a block diagram of the major components of the computer and indicates the main interconnections. All the aforementioned registers are connected in various ways to a conductor known as the " digit trunk ", and it is along this conductor that the numbers are transferred from one register to another in the form of trains of suitably timed electrical pulses.

III. The Number and Command Code

A single location in the stores contains a single word or datum consisting of 20 binary digits, the distribution of which represents a single number, command, or partitioned datum according to the manner in which the programme uses the word. The positions or periods of the digits of a 20-digit word are denoted by

symbols, p_1, p_2, . . ., p_{20}, and are written in the time sequence at which the digits appear on the digit trunk. Each digit held within any one register may possess a weight of either unity or zero, according as at the appropriate digit instant an electrical pulse respectively does or does not appear in the digit trunk. Single digits are frequently referred to by their corresponding period symbol, that is, a unit digit at the instant p_r is called a p_r digit, and the time of appearance of such a digit is the p_r period.

(a) Numbers

In the case of numbers a weighting is applied to the digits which relates one digit to that which immediately follows in the time sequence. Thus, a unit p_r digit is given the additional weighing of 2^{r-20}. Negative numbers are stored in their complementary form and the " binary point " is considered as lying between the digit periods of p_{19} and p_{20}. Hence a single number can take values equal to integral multiples of 2^{-19} from -1 to $1-2^{-19}$.

Such a convention considers numbers as fractions, but by suitable programme design it is possible to compute with numbers as though the binary point were between any pair of adjacent digits.

(b) Partitioned Data

Not all data held in the store are modulo two numbers. Some 20-digit data may be used by a programme according to any desired convention of partition. Thus a word may, in the course of a programme, be divided up into a number of separate components either related to one another or not, e.g. into the groups p_1-p_{10} and p_{11}-p_{20}. Examples of partitioned data are the two components of a complex number, numbers which bear " tags ", that is, special digits describing the meaning of the accompanying main group of digits, and numbers in floating form which possess variable indices.

All commands are a form of partitioned data.

(c) Commands

A single command is coded as a sequence of 20 binary digits. The address convention of partitioning adopted, however, affects the entire design of the computer and the method by which programmes are designed.

An operation defined by any one command is regarded as the transfer of the digital content of one register to a second register. Any arithmetical operation which takes place during a transfer depends on the manner in which the first or " source " register transmits, and the second or " destination " register receives. In this sense the command code system is of the two-address type.

The digits of any command are partitioned into three groups ; the first those in periods p_1-p_5, the second in p_6-p_{10}, and the third in p_{11}-p_{20} inclusive. The first group contains the code for the destination and the second the code for the source. The third, the largest group, is a " sub-address " and indicates one of the 1024 serially numbered locations in one or other of the stores if required. This last group may also contain any desired group of 10 binary

digits even if neither the source nor the destination calls upon the stores. The command code allows for a total of 32 sources and 32 destinations and stores of up to 1024 locations. It is believed that the C.S.I.R.O. Mark I is the only computer to use an address system of this type.

In cases of commands which refer to a location in the D register the digits in $p_{11}-p_{14}$ indicate which location is called for use irrespective of the digits in $p_{15}-p_{20}$.

The address groups and the digit periods are illustrated in Table 1.

TABLE 1

THE ADDRESS SYSTEM

	$p_{20}\, p_{19}\, p_{18} \cdots\cdots p_{11}$	$p_{10}\, p_9\, p_8\, p_7\, p_6$	$p_5\, p_4\, p_3\, p_2\, p_1$
Digit periods	$p_{20}\, p_{19}\, p_{18} \cdots\cdots p_{11}$	$p_{10}\, p_9\, p_8\, p_7\, p_6$	$p_5\, p_4\, p_3\, p_2\, p_1$
Address groups	←Store sub-address→	←—Source—→	← Destination →
Mod. 2 number weightings	Sign $2^{-1} \ldots\ldots\ldots 2^{-9}$	$2^{-10}\ldots\ldots\ldots$	$\ldots\ldots\ldots 2^{-19}$
Typical command $c(A) \rightarrow 16$	0 0 0 0 0 1 0 0 0 0	0 1 0 0 1	0 0 0 0 0
The number 4/5	0 1 1 0 0 1 1 0 0 1	1 0 0 1 1	0 0 1 1 0

IV. Function Gates

To every source code number there exists a " source function gate " which becomes activated and able to transmit a datum to the digit trunk, whenever the current command contains the appropriate number in the source group. Similarly every destination code number corresponds to a " destination function gate " which is activated according to the destination address of the current command and allows a datum to pass through it from the digit trunk. These are indicated in Figure 1.

V. The Operation Sequence

Commands to be performed in succession are normally stored in and withdrawn from successively numbered locations in the high speed store. For any one command to be satisfied four transfers must take place in the following sequence :

(1) The content of the sequence register is transmitted to the store control register, thus preparing the store to transmit out the next command.

(2) The store transmits the contents of the position indicated by the content of the store control register, i.e. the new command, to the interpreter register.

(3) The sub-address group ($p_{11}-p_{20}$ digits) of the current command is transmitted back from the interpreter register to the store control register, being substituted in place of the previous contents. This prepares the store for action if either of the addresses requires it.

(4) The appropriate source and destination function gates indicated by the addresses of the command are activated by suitable circuits connected to the interpreter register, thus allowing the desired transfer to take place.

This sequence of transfers, called the " computer routine " is invariable and is controlled by the " main control unit " shown in Figure 1, which provides the necessary electrical waveforms.

The content of the sequence register at stage 1 of the computer routine is known as the " control number ". The term " control at n " implies that the current command is selected from location n. Any change in the value of n is " shift of control ". Normally control shifts from n to $n+1$ during a single cycle of the computer routine. A single unit p_{11} is added to the control number in the sequence register during stage 3 of each computer routine cycle.

VI. The Source and Destination Codes

The following tables list the operations for which the various function gates are responsible. Table 2 lists the sources and Table 3 the destinations in the numerical order of the code from 0 to 31.

TABLE 2
SOURCE FUNCTION GATES

Code Number	Function	Symbol
0	Read out and hold the content of the high speed store location n ($0 < n < 1023$), indicated by the sub-address digits $p_{11}-p_{20}$	(n)
1	Read out the content of the input register (20 digits, p_1-p_{20}) ..	(I)
2	Read out the content of hand-set register No. 1 (20 digits, p_1-p_{20})	(N_1)
3	Read out the content of hand-set register No. 2 (20 digits, p_1-p_{20})	(N_2)
4	Read out the content of register A (20 digits, p_1-p_{20})	(A)
5	Read out the most significant digit of the content of register A (unit p_{20} if 1, zero if 0)	$p_{20}.(A)$ or $s.(A)$
6	Read out the content of register A divided by 2 (20 digits)	$\frac{1}{2}(A)$
7	Read out the content of register A multiplied by 2 (20 digits)	$2(A)$
8	Read out the least significant digit of the content of register A (unit p_1 if 1, zero if 0)	$p_1.(A)$ or $l.(A)$
9	Read out the content of register A and leave it cleared to zero	$c(A)$
10	If the content of register A is non-zero transmit a unit p_1 digit, otherwise transmit zero	$\tilde{z}(A)$
11	Read out the content of register B	(B)
12	Read out a unit p_1 digit if the most significant digit of the content of register B is unity, otherwise transmit zero	(R)
13	Read out the content of register B shifted to the right one place	$r(B)$
14	Read out the content of register C	(C)
15	Read out the most significant digit of the content of register C (unit p_{20} if 1, zero if 0)	$p_{20}.(C)$ or $s.(C)$
16	Read out the content of register C shifted one place to the right (zero in p_{20} position)	$r(C)$
17	Read out the content of the location in register D indicated by the number m ($0 < m < 16$), represented by the digits $p_{11}-p_{14}$	(D_m)
18	Read out the most significant digit in the location in register D indicated by the number m ($0 < m < 16$), represented by the digits $p_{11}-p_{14}$	$p_{20}.(D_m)$ or $s.(D_m)$

TABLE 2 (*Continued*)

SOURCE FUNCTION GATES

Code Number	Function	Symbol
19	Read out the content of the location in register D indicated by the number m $(0 \leqslant m < 16)$, represented by the digits p_{11}–p_{14}, shifted one place to the right	$r(D_m)$
20	Read out zero	(Z)
21	Read out the 10-digit content of register H in the position group p_1–p_{10}	(H_l)
22	Read out the 10-digit content of register H in the position group p_{11}–p_{20}	(H_u)
23	Read out the 10-digit content of the sequence register in the position group p_{11}–p_{20}	(S)
24	Read out a single unit digit in the p_{11} position	p_{11}
25	Read out a single unit digit in the p_1 position	p_1
26	Read out from the interpreter register the content of the digits of the current command held in positions p_{11}–p_{20}	(K)
27	Read out the content of the low speed store No. 1 from the location n $(0 \leqslant n < 1024)$, indicated by the 10-digit sub-address of the current command	(n_1)
28	Read out the content of the low speed store No. 2 from the location n $(0 \leqslant n < 1024)$, indicated by the 10-digit sub-address of the current command	(n_2)
29	Read out the content of the low speed store No. 3 from the location n $(0 \leqslant n < 1024)$, indicated by the 10-digit sub-address of the current command	(n_3)
30	Read out the content of the low speed store No. 4 from the location n $(0 \leqslant n < 1024)$, indicated by the 10-digit sub-address of the current command	(n_4)
31	Read out a unit digit in the position p_{20}	p_{20}

Most of the entries in Tables 2 and 3 are self-explanatory. The following notes will assist in defining more clearly the action of some of the functions.

Not all registers are associated with both source and destination gates; for instance, the output register possesses only destination gates, whilst the various constants have only source gates.

The output from any source gate may be transferred to any destination gate which may be consistently coded. Thus, to transfer the content of location n in the high speed store to location m $(n \neq m)$ in the high speed store is an invalid transfer and cannot be coded, except in the case $n = m$ when the transfer is possible. Similarly a transfer from a store location $16r + p$ $(0 \leqslant p < 16)$ to location q $(0 \leqslant q < 16)$ of register D is possible only if $p = q$. Similar remarks apply to the low speed store.

The input register possesses 20-digit hand-setting facilities like those of N_1 and N_2 and the entire 20 digits are transferred if called as a source.

The destination I_t affects the mode of reading data from the medium, in this case punched paper tape, from which data are placed into the input register. In the " read binary " condition the row of digits represented by the configuration of holes or the current column of the tape is accepted directly into the register. In the " decimal read " state the configuration of digits is subjected to some transformation upon entering the input register. The effect in this state is that of interpreting tape punched with one hole per row in a 1 to 10 code for decimal digits and transferring each digit to the input register as the equivalent binary tetrad. This is more fully described later.

In the right-shift sources of registers B, C, and D, i.e. $r(B)$, $r(C)$, and $r(D)$, the lowest digit (at p_1) is omitted from transfer ; a unit digit in the p_r period is transmitted as a unit p_{r-1} and the digit transmitted in the p_{20} position is zero irrespective of the digit in the p_{20} period held in the register.

In the case of multiplication the modulo 2 convention is adopted and the product of the content of register C and the number entering register B, both stored modulo 2, is added into the combined registers A and B as though the binary point of the product lies between the p_{20} and p_{19} periods in register A. The original multiplier is lost in the process of multiplication and is replaced by the less significant 19 digits of the product. The multiplicand is retained in register C. In cases of the multiplier or multiplicand being negative (i.e. possessing unit p_{20} digits) suitable corrections are made to the product to provide a negative product in complementary form.

A loudspeaker has been found useful for providing audible signals which help the operator to follow the course of a calculation. It is selected by destination number 10.

VII. NOTATION

A special notation is adopted for writing programmes. Each command is represented by at least two symbols, one for the source and one for the destination. The symbols are those indicated in Tables 2 and 3. Usually these symbols involve a letter denoting the register involved and an accompanying symbol to indicate the particular function of the function gate selected.

The source symbol is written to the left of a rightward pointing arrow with a lower half-head only, i.e. $\rightharpoonup$, to the right of which the destination symbol is written. To aid the eye the functional component of the destination symbol is moved to a point over the arrow. This symbol does not exist on standard typewriters, but this is not a great inconvenience since programmes are finally typed on standard forms which have the arrow printed upon them.

The content of a register is implied by surrounding the letter representing the register referred to by round brackets, thus (A) means the numerical content of register A and is the symbol shown in Table 2. The bracket notation is used in indicating some source functions. In cases of the remaining sources the brackets are not necessary as, for instance, when a constant digit or set of digits is transmitted, e.g. sources of p_1, p_{11}, and p_{20} units. Further, if the source is the interpreter register, it is convenient to write the actual number transferred out in place of the source symbol omitting brackets, the interpreter register being

<div align="center">

TABLE 3

DESTINATION FUNCTION GATES

</div>

Code Number	Function	Symbol
0	Read and substitute into the high speed store in location n $(0 \leqslant n \leqslant 1023)$, indicated by the digits $p_{11}-p_{20}$ of the current command	n
1	If at least one unit received, change the mode of input reading from decimal code to straight binary or vice versa, otherwise no change	I_t
2	Read into the output register the digits received in positions $p_{11}-p_{15}$ and print the corresponding character	O_t
3	Read into the output register the digits received in positions p_1-p_5 and punch onto tape	O_p
4	Substitute into the register A (20 digits)	A
5	Add into the content of register A and hold the sum ..	$+A$
6	Subtract from the content of register A and hold the difference	$-A$
7	Replace the content of register A by the digit by digit product of its content and the entering digits (e.g. conjunction)	$.A$
8	Replace the content of register A by the digit by digit logical sum of its content and the entering digits (e.g. disjunction)	$\vee A$
9	Replace the content of register A by a unit digit wherever the content differs digit by digit from the digits entering	$\sim A$
10	Transfer the entering digit train into the loudspeaker ..	P
11	Substitute into register B	B
12	Read into register B, form the product of the content of B and register C (modulo 2) and add into the content of register A, retaining the lowest 19 digits of the product in B with a zero in the p_1 position of B	$\times B$
13	Only if a unit p_{20} is received shift the content of registers A and B one place to the left, the digit in p_{20} position of B shifting to the p_1 position of A	L
14	Substitute into register C	C
15	Add into the content of register C and hold the sum	$+C$
16	Subtract from the content of register C and hold the difference..	$-C$
17	Substitute into the location m $(0 \leqslant m < 16)$ of register D, indicated by the $p_{11}-p_{14}$ digits of the current command	D_m
18	Add into the content of location m $(0 \leqslant m < 16)$ of register D, indicated by the $p_{11}-p_{14}$ digits of the current command, and hold the sum	$+D_m$
19	Subtract from the content of location m $(0 \leqslant m < 16)$ of register D, indicated by the $p_{11}-p_{14}$ digits of the current command, and hold the difference	$-D_m$
20	Null	Z
21	Substitute into register H the digits from group p_1-p_{10} of the entering number	H_l
22	Substitute into register H the digits from the group $p_{11}-p_{20}$ of the entering number	H_u
23	Substitute into the sequence register the digits entering in the group $p_{11}-p_{20}$	S

TABLE 3 (Continued)
DESTINATION FUNCTION GATES

Code Number	Function	Symbol
24	Count the number of digits entering into the content of the sequence register (each unit digit has unit weighting)	cS
25	Add into the content of the sequence register the digits entering in the group p_{11}–p_{20} and hold the sum	$+S$
26	Substitute the p_{11}–p_{20} digit group of the interpreter register by the p_{11}–p_{20} of digits entering. Hold the content until the next command enters when the corresponding group will be added in	$+K$
27	Substitute in the low speed store No. 1 into the location n ($0 \leqslant n < 1024$), indicated by the p_{11}–p_{20} group of the current command	n_1
28	Substitute in the low speed store No. 2 into the location n ($0 \leqslant n < 1024$), indicated by the p_{11}–p_{20} group of the current command	n_2
29	Substitute in the low speed store No. 3 into the location n ($0 \leqslant n < 1024$), indicated by the p_{11}–p_{20} group of the current command	n_3
30	Substitute in the low speed store No. 4 into the location n ($0 \leqslant n < 1024$), indicated by the p_{11}–p_{20} group of the current command	n_4
31	If one or more unit digits received, stop the sequence ; do not proceed to the next command	T

then understood as the source. The symbol (K) for the interpreter source is used only in the early design stage of a programme when the sub-address number may be unknown until a later stage in the design when the symbol is replaced by the actual number.

The following examples illustrate the notation :

(i) $(C) \rightarrow A$: Read out the content of register C and substitute it into register A.

(ii) $(57) \overset{+}{\rightarrow} A$: Read out the content of store position 57 and add it into register A.

(iii) $57 \rightarrow S$: Read from the interpreter the digits in positions p_{11}–p_{20}, i.e. 57 p_{11}, and substitute them into the sequence register. Note that (K) has been replaced by the number 57 with no brackets.

(iv) $p_{11} \overset{+}{\rightarrow} D_4$: Read a p_{11} unit from the p_{11} source and add it into location No. 4 of the register D.

Commands comprising a programme are written one under the other in the order in which they are entered into the store. Usually commands are placed into successive locations. To the left of each command symbol a serial number

or the store location into which the command is to be transferred upon entry into the computer is written.   To the right of the list of commands the effect or action of commands or groups of commands is stated and any additional notes are made.

Two vertical lines placed to the left and three to the right of the list of commands provide one column between the location number and the commands and two adjacent spaces between the commands and the notes on the right.   An example of the layout will be found in Table 4.

Commands which change during the course of the operation of the programme are surrounded by square brackets, [], whilst commands which are not actually used by the machine directly as such, but which are used as quantities from which other commands may be built, are surrounded by angle brackets, $<>$, and are called " pseudo- " or " latent commands ".   When written symbols are to be grouped together they are overlined, e.g. $\overline{n+p+r}$.

It is sometimes necessary in descriptions of programmes to refer to the content of a register at the completion of a particular command.   Such a quantity is indicated by a prime.   Thus $(A)'=x$ written to the right of a command means " the content of register A at the completion of the command is x ".

During the normal course of the operation of the machine through the programme, commands are adopted from successively numbered locations in a uniform sequence.   Such a sequence is frequently broken, the next command being adopted from a location other than the next in numerical order.   In such cases the serial number of the command next selected is indicated to the right of the current command in the second or rightmost of the blank columns formed by the vertical lines and is enclosed in square brackets.

A list of commands is sometimes broken into convenient groups of commands by underlining the last of such a group.

It is not always possible to denote quantities entered into the computer in the form of commands or pseudo-commands.   Special numbers may be written in the decimal scale during the initial stages of design of a programme and later converted into a form which is suitable for recording onto the input medium, the paper tape.   Such special symbols which the programmer may use should be placed in a position which would otherwise be occupied by a command symbol.

Sometimes, but only rarely, the sub-address may not be entered directly into the command symbols constructed from Tables 2 and 3.   In such cases the sub-address number is written to the left of the command symbol, e.g. 57 ; $(C) \rightarrow A$ has the effect of the command $(C) \rightarrow A$, but is also the number $57 \times 2^{-9}+46 \times 2^{-19}$, and may be used as such.   The number 57 affects only the time in the computer routine at which the transfer takes place.

It is sometimes convenient to use the notation for digit-periods to represent the important unit occurring in a sub-address.   Thus a number, say n, representing a location, is normally stored in the $p_{11}-p_{20}$ group only and may be written in the command in the form np_{11}, but usually the p_{11} is implied.   In other cases the notation is extended to cover digits in which the p_{11} unit is not the important unit, e.g. mp_{15}.

VIII. Programme Devices

Not all the 32×32 possible combinations of source and destination transfers are useful and some inefficiency in the use of digit space in a programme must be accepted. However, compared with a straight one-address system considerable extra flexibility and some saving in the total number of commands in programmes which serve similar purposes is attained.

Certain elements in the design of the machine distinguish the method adopted for programme design from those used for other computers. These are the main factors; first, the computer possesses more than one accumulating register. In fact A and C and all 16 locations in D are capable of the elementary arithmetical functions, although register A has extra facilities and in practice becomes the main accumulator and is used in preference to other registers when available. The accumulating registers and the registers B and H may be used as " working space " during a calculation and a considerable saving of commands is attained, compared with an equivalent programme designed for a one-address system, by calling transfers between these registers without reference to the main store. Further, these registers are frequently used at one stage of a programme for the assembly of data to be operated upon by parts of the programme immediately to follow. Thus, for instance, prior to a programmed division the dividend is transferred to register A and the divisor to C and the division which follows uses only one further register, B, and the quotient appears in register A.

Facilities are available for making discrimination, that is, choosing alternative paths through a programme according as a condition is or is not satisfied by the content of a particular register. Discrimination is usually made according to the state of the sign of a register content, that is, the p_{20} digit of the content in question is either unity for negative content or zero for positive content including zero. Sign conditions of (A), (C), and (D_m) $(0 \leqslant m < 16)$ may be detected by the sources $s(A)$, $s(C)$, and $s(D_m)$, and also on the content of B by the source (R). The evenness or oddness of (A) may be detected by the source $p_1.(A)$. Digits from these gates may be transferred to any desired destination but the most common ones are described below.

Breaks in the normal sequential course through a programme of commands are frequent. Commands which cause these breaks involve sequence register destinations. The choice of alternative paths through a programme can be made via the destination cS. Usually only one digit is allowed to enter this gate. Thus, the command

$$s(A) \xrightarrow{c} S$$

from location n in the store will cause the next command to be selected from the position $n+1$ immediately following if the sign digit of (A) is zero, or from the position $n+2$ if unity.

Sometimes, however, numbers of digits may be counted in the same transfer, thus making possible a choice between a number of alternative operations to follow. Thus the command in position n

$$(A) \xrightarrow{c} S$$

will, if r unit digits are contained in (A), cause the next command to be selected from position $n+r+1$. Such a break in the uniform progress through a programme is called a " conditional shift of control ".

The substitution of an entirely fresh number into the sequence register via the destination S causes an " absolute shift of control ". Thus the command in position n

$$(A) \rightarrow S$$

replaces (S) by the 10 most significant digits of (A), i.e. those in the period $p_{11}-p_{20}$, neglecting any digits in the group of lower significance, p_1-p_{10}. Thus, if (A) is the partitioned datum

$$(A)=qp_{11}+rp_1, \qquad 0 \leqslant q < 1024, \ 0 \leqslant r < 1024,$$

(S) becomes qp_{11} and the next command is withdrawn from position q, that is,

$$(S)'=q.$$

A " relative shift of control " can be attained via the destination $+S$. Thus the command in location n

$$(A) \overset{+}{\rightarrow} S,$$

where again $(A)=qp_{11}+rp_1$, causes control to be shifted " forward " to $n+q+1$, that is,

$$(S)'=n+q+1.$$

The extra unit is introduced because the addition of a p_{11} unit, which controls the normal advance of (S) from n to $n+1$, is not removed. Normally such an operation represents an increase in the actual value of (S). Since, however, (S) is a number stored modulo 1024 in terms of p_{11} as a unit, the addition of the complement of q, i.e. $\overline{1024-q} \ p_{11}$, will cause $(S)'$ to be $n-q+1$ to which control will be shifted " backwards ".

The sequence register may also be called as a source. Thus a command in location n

$$(S) \rightarrow A$$

substitutes $\overline{n+1} \ p_{11}$ into register A. Such an operation is usually followed by a command causing (S) to change, say, to q, a point in the programme which might be reached by more than one route. As the programme proceeds from location q a reference to the route along which q was reached is stored for use as (A).

Many of the functions just described are used in conjunction with the interpreter register as source. In this case the digits in positions $p_{11}-p_{20}$ of the current command are available for use. This is so whenever the sub-address of the current command does not refer to a position in the stores or register D. Thus, in accordance with the notation previously stated, the command in location n

$$57 \rightarrow S$$

would cause the number $57p_{11}$ to be transmitted from the interpreter and to be substituted into the sequence register. Control would be shifted to 57.

The command in location n

$$57 \overset{+}{\to} S$$

would cause control to be shifted forward to $57+n+1$.

The notation is important here ; the round bracket is not present and the interpreter is implied as the source. By contrast, the command if written as

$$(57) \overset{+}{\to} S$$

would cause the most significant group of 10 digits of the content of store position 57 to add into (S).

In conjunction with the H register current commands may be used to store useful constants. To form a full 20-digit datum from a pair of 10-digit groups available in the current commands the H register is used for shifting one group from the $p_{11}-p_{20}$ period to the p_1-p_{10} period. A typical set of commands would be :

Location of Command	Command
n	$r \to A$
$n+1$	$s \to H_u$
$n+2$	$(H_l) \overset{+}{\to} A,$

where the first places rp_{11} into A, the second places sp_{11} into H, which is read out by the third as sp_1 and added into A. Thus $(A)'=rp_{11}+sp_1$ or $r2^{-9}+s2^{19}$ numerically.

The basic constants, the p_{20}, p_{11}, and p_1 unit digits, are used largely as follows : p_{20} for the constant -1 and sign reversals and modulo 2 counting, p_{11} for unit counts, in particular for changes by unity to the sub-address of commands changed by the programme, and p_1 for unit counts to complete 20-digit numbers. etc.

The operation of the computer may be stopped either conditionally or unconditionally. Thus the command

$$s(A) \to T$$

will stop the sequence of operation if (A) contains a p_{20} unit, otherwise not. An unconditional stop is obtained by a command such as

$$p_{20} \to T \text{ or } p_{11} \to T.$$

Manual control may be applied in any case by use of the switches in registers N_1 and N_2. Thus

$$(N_1) \to T$$

will stop the operation only when at least one switch of N_1 is depressed.

Registers N_1 and N_2 are frequently used to hold parameters which may be varied at will by the operator, e.g. parameters which control the rate of convergence of an iterative process.

A register may read out of one of its source gates and receive into one of its destination gates at the same time. Thus the command

$$(A) \overset{-}{\to} A$$

causes (A) to be read out and subtract into itself so that $(A)'$ is zero. Alternatively,

$$c(A) \overset{-}{\rightarrow} A$$

causes (A) to be read out and at the same time cleared to zero and the original content transmitted is received and subtracted from zero. The effect is to replace (A) by $-(A)$.

Simple multiples of (A) may be obtained by the sources $\frac{1}{2}(A)$ and $2(A)$ functions. Thus

$$2(A) \overset{\scriptscriptstyle \tau}{\rightarrow} A$$

gives $(A)'=3(A)$ and so on.

Operations of the form $(D_m) \overset{+}{\rightarrow} D_m$ and $(D_m) \overset{-}{\rightarrow} D_m$, i.e. doubling and clearing the mth location in D are permissible.

Multiplication by a quantity already stored in register B may be caused by the command

$$(B) \overset{\times}{\rightarrow} B,$$

which causes (B) to be read out and to enter the multiplying gate and re-enter B and so start the multiplication process.

The command $c(A) \overset{\times}{\rightarrow} B$ is frequently useful in forming continued products when (A) is already the product of a previous multiplication.

Occasions arise when " time wasting " must take place as when a command is removed from the programme leaving a " free space ". Such spaces may be filled with what may be called " null commands ", either $(A) \rightarrow A$, which reads (A) back to the same place, or by any " store to store " command like $(n) \rightarrow n$.

It will be noticed that the code representing such a command is equivalent to the number $np_{11}+0p_1$, so that a blank space $0p_{11}+0p_1$ causes the content of store location 0 to be resubstituted into the same location. Thus, blank spaces, if adopted as programme commands, affect no registers or stores.

As a simple example of the construction of a programme consider the commands which will cause the product of the modulus of the content of store location 47 to be multiplied by 2^{-4}, rounded off to 19 binary places, transferred to location 12, and control shifted to 132. This set of commands, listed in Table 4 in the standard manner, is entered at the leading command placed in location m.

The action caused by these commands may be followed from the description in the right-hand column. The command in $m+1$ tests the sign of (A), and causes a conditional shift of control, and, if positive, (A) is replaced by its complement *twice*, thus leaving it positive. This set of commands is not unique and other arrangements can be contrived which will provide the same final effect.

IX. Input and Output

Programmes, following the initial design stages, are written in the symbolic form in the order in which they are to be provided to the computer. Thence they may be transcribed onto the input medium. Both commands and numerical data are supplied to the machine on punched paper tape onto which up to 12

holes per row can be recorded. Single commands when punched onto tape
normally occupy two rows consisting of two groups of 10 binary digits each
corresponding to 10 of the 12 punching positions in a row. The remaining two
positions are reserved for special controlling purposes. One column is punched
with the binary representation of the sub-address intended to be placed in periods
p_{11}–p_{20} and the following column with the other addresses to enter p_1–p_{10}. The
latter is supplied with a hole in the 11th or X position, and distinguishes p_1–p_{10}
rows from p_{11}–p_{20} rows.

When the machine is switched on all registers and the store are cleared to
zero and a group of 20 commands is transferred to the first 20 locations in the
store (0–19) via the fixed wiring on a set of rotary stepping switches. This set

TABLE 4

A " ROUNDED MODULUS " ROUTINE

Location	Command		Action
$m+0$	$(47) \rightarrow A$		$(A)' = (47)$
$m+1$	$s(A) \overset{c}{\rightarrow} S$		Tests $(A) \gtrless 0$
$m+2$	$c(A) \overset{-}{\rightarrow} A$		Skipped if $(A) < 0$, if $(A) \geqslant 0$, $(A)' = -A$
$m+3$	$c(A) \overset{-}{\rightarrow} A$		If at $m+1$ $(A) \leqslant 0$, then $(A)' = -(A)$, otherwise $(A)' = (A)$
$m+4$	$32 \rightarrow C$		Sets a p_{16} unit to C from interpreter
$m+5$	$c(A) \overset{\times}{\rightarrow} B$		Forms $\mid (47) \mid \times 2^{-4} = (A)'$
$m+6$	$(R) \overset{+}{\rightarrow} A$		Round-off to 19 binary digits
$m+7$	$(A) \rightarrow 12$		$(12)' = \mid (47) \mid \times 2^{-4}$
$m+8$	$132 \rightarrow S$	[132]	Shifts control to 132

of commands is sufficient to assemble the digits from pairs of adjacent rows of
the input tape into full 20-digit data and to store them in successive positions in
the store immediately following those already filled. This set of commands is
called the " primary routine ".

In cases of commands which have zero as the sub-address the first of the
pairs of rows may be omitted. In all commands a punch in the X position is
provided on the column containing the source and destination addresses so that
the total number of such holes is equal to the number of commands punched.

A special keyboard of 32 keys is provided for transcribing programmes
onto tape. Each key corresponds to a number in the scale 32, to a source, and to
a destination. The number and the corresponding address symbols of Tables 2
and 3 are indicated on each key. Two keys are depressed in succession before
one of a set of " punch keys " is depressed which initiate the punching. Thus
the sub-address is punched by depressing first the key m, say, followed by n and
the punch key is depressed when the 10 binary digits are punched in order on the

tape row currently under the tape punches.   The most significant digit is punched
in the first position, the others follow in order to the lowest in the 10th position.
The integer thus punched will be : $32m - n$.   Following the punching of a row
the tape is automatically moved forward to the next row.   The source key is
depressed followed by the destination key and the " X-punch " key is depressed.
The source and destination code is then punched together with punch in the X
position.   There is no need to remember the actual code numbers of the addresses.
The keying for a typical command is therefore :

$$m, \; n, \; \text{punch} \; ; \; \text{source}, \; \text{destination}, \; X \; \text{punch}.$$

As an example, the programme at the end of the previous section would be
punched by depressing keys in the sequence shown in Table 5.   In this table
semicolons imply the depression of the punch key, whilst the colons imply the
depression of X-punch key.

<div align="center">

TABLE 5

PROGRAMME ENCODING METHOD

Example shown for the routine in Table 4

</div>

Depression of Keys	Corresponding Written Command
1 , 15 ; (0) , A :	$(47) \rightarrow A$
$s(A)$, cS :	$s(A) \xrightarrow{c} S$
$c(A)$, $-A$:	$c(A) \xrightarrow{-} A$
$c(A)$, $-A$:	$c(A) \xrightarrow{-} A$
1 , 0 ; (K) , C :	$32 \rightarrow C$
$c(A)$, $\times B$:	$c(A) \xrightarrow{\times} B$
(R) , $+A$:	$(R) \xrightarrow{+} A$
0 , 12 ; (A) , 0 :	$(A) \rightarrow 12$
4 , 4 ; (K) , S :	$132 \rightarrow S$

All the symbols in Table 5 are those listed in Tables 2 and 3 and almost all
are the same as are used in writing programmes onto paper.   The exception,
symbol (K), exists on the key representing the interpreter as source and is
implied in the written commands.   With this latter provision the list of written
commands may be translated into keying operations with no rewriting and very
little mental effort.

In cases where a 20-digit binary number to be entered into a programme is
not a command and is not necessarily in the form of a pseudo-command, the
number, multiplied by 2^{19}, must first be converted to the scale of 32 and keyed
in the same manner as a normal command.   Such data are constants like e^{-1},
$\pi \cdot 4$, $2^{-\frac{1}{2}}$, etc.   Thus, for instance, in the final steps of compiling the programme
$\pi/4$ would be entered by keying the sequence :

$$12, \; 18, \; \text{punch} \; ; \; 3, \; 31, \; X \; \text{punch},$$

since $0 \cdot 7853982 / 2^{19} = 12, \; 18, \; 3, \; 31$ to the nearest unit in scale of 32.

Normally very few such cases occur, as decimal data in large quantities are normally punched in one of two ways ; each decimal digit is punched either in a decimal code similar to that used on cards of the Hollerith type, i.e. the 1 in 10 position code, or as an equivalent binary tetrad. When using either of these codes part of a programme previously entered into the store must be designed so as to assemble the digits transferred from the punch holes into the binary representation needed later by the programme.

Results are normally page printed or punched onto tape. The computer is provided with a suitable printing or output punching programme, which controls the transformation of the results from binary form as stored internally to the corresponding decimal digits, which are printed or punched in descending order of significance. The printer is provided with a full range of letters, brackets, stops, and positive and negative signs, which may be used by suitable programmes.

Decimal results may be punched in the form of binary tetrads or as single holes in the 1 in 10 code system. This allows punched results to be reinserted into the computer during a later calculation.

X. REFERENCES

GILL, S. (1951).—The diagnosis of mistakes in programmes in the EDSAC. *Proc. Roy. Soc.* A **206** : 538.

PEARCEY, T., and HILL, G. W. (1953).—Programme design for the C.S.I.R.O. Mark I Computer. II. Programme techniques. *Aust. J. Phys.* **6** : 335.

WHEELER, D. J. (1950).—Programme organization and initial orders for the EDSAC. *Proc. Roy. Soc.* A **202** : 573.

WILKES, M. V., WHEELER, D. J., and GILL, S. (1951).—'' The Preparation of Programmes for an Electronic Digital Computer.'' (Addison-Wesley Press Inc.: Cambridge, Mass.)

PROGRAMME DESIGN FOR THE C.S.I.R.O. MARK I COMPUTER

II. PROGRAMME TECHNIQUES

By T. Pearcey* and G. W. Hill*

[*Manuscript received December 5, 1952*]

Summary

The general structure of typical programmes is considered in relation to the structure of the computer, in particular to its facilities which tend to render programmes invariant with regard to their position in the store. Full-scale programmes are constructed by piecing together by a master programme a suitable selection of items from a library of " standard routines ". The library contains " sub-routines " which are completely self-contained and require no additional information for their operation and are invariant with respect to their position in the store, and " routines " which are not so invariant, and frequently require additional data to be provided during their entry into the store. The entry of special data for routines and the simplification of the construction of the master programme are facilitated by a special routine used whilst the entire programme is being entered into the store.

I. Introduction

Part I (Pearcey and Hill 1953) dealt largely with the preliminaries to the design of full-scale programmes. It described the computer, its address system, and the system of notation used for designing and recording programmes on paper prior to the stage of translating them onto a medium for machine input. The written programme is recorded onto punched tape with the aid of a specially modified teleprinter keyboard. The written programme permits direct translation onto tape without further writing or special effort.

Part II includes a general discussion of the nature of programmes, the main features being illustrated with regard to the use of the present machine. Basic groups of commands or " sub-routines " are selected from a library of such groups and only the connections between them, together with any special operations, are included in the part of the programme specially designed for the current problem. Commands may be modified upon entering the computer in a number of ways which involve use of a special routine which is entered initially ; this routine is described in detail.

II. The Elements of a Programme

The ability of a computer to change the commands held in the store in an organized manner permits a great saving in the effort needed for programme design and also in the number of commands required to perform a required calculation. In practice only a small proportion of the commands of a

---

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

programme is actually devoted to making the calculation; the greater proportion organizes the course of calculation and frequently involves continued adjustment of some commands.

A typical example of a programme having the normal structure is given in Table 1. The construction of the table may be compared with those for a similar example given for the EDSAC (Wilkes, Wheeler, and Gill 1951).

Suppose it is required to find the sum of the content of all storage locations from 100 to 149 inclusive, to store the result in register A, and then stop. The notation described in Part I is used in Table 1, and the programme is entered at the first command in the list placed at location m.

TABLE 1

TYPICAL PROGRAMME STRUCTURE

Location of Command	Command			Action
m	$(A) \overset{-}{\to} A$			Clears A, $(A)'=0$
$m+1$	$(149) \overset{+}{\to} A$			Addition starts with (149)
$m+2$	$(m+1) \to C$			Reduces $(m+1)$ to $(149-r) \overset{+}{\to} A$;
$m+3$	$p_{11} \overset{=}{\to} C$			$r=0, 1, 2, \ldots$
$m+4$	$(C) \to m+1$			
$m+5$	$100 \overset{-}{\to} C$			Reduces (C) by $100\, p_{11}$
$m+6$	$s(C) \overset{c}{\to} S$			Tests sign of (C)
$m+7$	$m+1 \to S$		$[m+1]$	If <50 addition, i.e. $r<50$
$m+8$	$p_{20} \to T$			$r=50$; additions complete; stop

The programme possesses the following features which illustrate the normal structure of most programmes:

(1) The programme is partly repeated, being re-traversed from the command in $m+1$ to that in $m+7$.

(2) During each traverse of the repeated commands only one command, that in $m+1$, is directly devoted to the calculation which is the objective of the programme.

(3) The command performing the desired arithmetical function is changed during each traverse by the computer so that in the following traverse it refers to a different, although adjacent, storage location.

(4) Three of the six repeated commands adjust the command in location $m+1$, i.e. those in $m+2$, $m+3$, and $m+4$.

(5) A count is made of the number of repetitions, and is tested during each traverse to decide whether or not the full number of additions has been made.

(6) If the summation is incomplete, control is shifted " backwards " to $m+1$, otherwise a " stop " command, as in $m+8$, is reached.

(7) One command is not repeated, i.e. that at m. This is devoted to the preparatory function of clearing register A.

In a complete programme the command in location $m+8$ would be replaced by one which would cause a shift of control to continue with further calculation.

From Table 1 it is seen that a programme consists of :

(a) A preparatory group of commands.

(b) A group performing the functions which are the objective of the programme.

(c) A group of commands which adjusts the programme for successive immediate repetitions.

(d) A group keeping a tally of the number of repetitions.

(e) A test on the tally to decide if the repetition shall cease and another part of the programme be adopted.

In some cases these sections of a simple programme may not be distinct and may even be absent. Thus, in an iterative process, as distinct from a repetitive process illustrated by Table 1, the groups (c) and (d) may be omitted, whilst (e) is present in the form of a test at each stage on the result from the group (b).

In the example given, (a) corresponds to the command in m, (b) to that in $m+1$, whilst $m+2$, $m+3$, and $m+4$ correspond to (c). The command which, at each stage, appears in $m+1$ is held in register C and is used as the " counter " for the repetitions.

Normally, programmes will contain a number of groups of commands similar in structure to that of Table 1. In fact, the main controlling programme will frequently be of this typical type in which its component parts are themselves smaller groups possessing similar structure and so on.

III. Standard Routines

Construction of each programme *ab initio* upon its individual merits would become very tedious. To save much of the labour of programme design as much use as possible is made of information accumulated and made available from previous programmes in the form of " standard routines " or groups of commands.

It frequently occurs that computations involve somewhat complex operations of standard types, such as evaluation of particular functions like exponential or circular functions, interpolation of a function table, numerical quadrature and so on. Each such operation requires its special programme which, if suitably designed, may be used as a standard routine for use in other calculations which require these types of operations.

Operations such as division, taking square roots, and even reading and recording fresh information and converting data to and from decimal scale may be so standardized. Such standard routines have the effect of providing the computer with additional and more complicated functions. Thus, for instance, an inversion routine may be considered as having the logical meaning, "substitute into register A the reciprocal of its present content ".

Two factors are necessary for successful use of standard routines in construction of programmes. Firstly, standard routines should be useful over as wide a

field as possible, that is, routines should not be too specialized otherwise the list of standard routines eventually recorded and stored becomes overwhelming and confusing. Secondly, there must be a simple means of incorporating them into programmes.

Standard routines frequently, although not always, possess a structure like that of Table 1 and may frequently involve more than one loop. For greatest usefulness it should be possible to place a routine in any part of the store without complication to the organization of the remainder of the programme. It should preferably be invariant with regard to its place in the store. Occasions will arise, particularly in use of routines for specialized use, when a routine must be restricted to a special position in the store and will not be invariant.

Further, a routine may require different data when used in a different context and may do so even when used more than once by the same programme. Such data are commonly called " parameters " and are supplied to the programme during its insertion into the machine, as will be described later.

Quantities used by a routine which differ every time the routine is entered are stored in standard positions, e.g. registers A, B, C, H, and D, just before the entry into the routine. These quantities are called " variables ".

IV. POSITION INVARIANCE

A special class of standard routines, which includes a large proportion of the most useful and frequently used routines, can be made invariant with regard to store position by use of the normal functions of the machine. The principal device used for this is the command of the type

$$n \overset{+}{\underset{-}{\to}} S,$$

which causes a relative shift of control by $n+1$ locations. In the example of Table 1 the command $m+1 \to S$ could be replaced by $-7 \overset{+}{\underset{-}{\to}} S$ (where -7 is considered modulo 1024). Within any routine the various values of n defining internal relative shifts of control are independent of the position of the routine in the store.

Numbers such as working constants, e.g. e^{-1}, $\pi/4$, etc., could be stored at the end of the routine by which they are used, that is in locations depending on the position of the routine in the store. This location reference can be` avoided by storing such constants in two parts within two successive commands. Thus the sequence of commands

$$n \to A,$$
$$m \to H_u,$$
$$(H_l) \overset{+}{\underset{-}{\to}} A,$$

places $(n 2^{10} + m) 2^{-19}$ in register A. Quantities may be similarly assembled into register C, but not in D, since the digit space occupied by n includes also the space specifying the location of D and may be incompatible.

By means of the further device of initially storing variables in various standard registers most of the simpler routines can be made invariant with regard

to store position.   Output routines and routines for division, square root, and simple transcendental functions are members of this class and are called " sub-routines ".   Those which require the use of parameters are called simply " routines ".

Not all functions which would help in rendering routines invariant have been included as functions of the machine.   Another would be a function which would allow a " transfer to the store location now indicated by the sequence register " to take place.   By this means direct reference to positions within the routine could be made, so that a command constructed with the aid of a position reference value drawn from the sequence register could be adopted for immediate use.   Such a procedure would involve a considerable number of additional commands.

V. Parameters

The use of standard routines may be greatly increased by use of parameters which may be given different values in different programmes.   Sub-routines do not possess such parameters.   Standard routines which possess internal references and cannot *conveniently* be made invariant may be provided with the store position of a particular command contained within it, usually its leading command, as one of its parameters.

Consider the example listed in Table 1.   The values depending upon the particular context are (a) 149, the first location summed, (b) 100, the last location summed, and (c) m, the location of the leading command of the routine.   There are three parameters, only one of which is determined by the way in which the routine is entered into the computer.

At this stage it is essential to distinguish between commands as they appear in the store after entry and the way in which the input tape is punched. Commands $m+1$, $m+2$, $m+4$, $m+5$, and $m+7$ may be entered in a special way dependent upon an extra character punched so as to distinguish them from the remaining commands.   This character may precede or follow the command affected and cause the value of the required parameter to be added to the command actually punched.   This may be done by use of a specially designed programme which must be placed in the store before the entry into the store of the programme.   If α, β, γ, and δ denote special characters, associated each with a parameter and with the commands of Table 1 into which the parameters are introduced, the programme designed for punching on the tape becomes that shown in Table 2.

The quantities entered into the second and sixth commands by β and γ must previously have been entered into the store.   The character α will have the effect of adding the location of the leading command of the routine to the sub-address immediately following.   This location number will have been stored upon reading the character δ at the head of the routine.

The details of the procedure whereby these operations are performed is described later.

The use of parameters, although somewhat complicating the entry of routines into the store, greatly increases their flexibility and renders them invariant with respect to position in the store.

VI. VARIABLES

Variables are normally held in the main arithmetical registers prior to entry by the programme into the routine or sub-routine which uses them. Thus the value x, of which the square root is required, is placed in register A before entry into a " square root routine ". Such a routine replaces x by $\sqrt{x}$.

Consider in particular the adaptation of the routine of Table 1 to the use of variables in place of parameters. Suppose, for instance, 149 be stored in A and 100 in H. Additional commands placed ahead of the command $(A)\overset{\cdot}{\to}A$ place the appropriate values for later use by the routine.

TABLE 2

INCORPORATION OF PARAMETERS

Example from Table 1

Relative Location of Command		Commands as Punched on Tape			Action of Additional Symbol
0	δ ;	$(A)\overset{\cdot}{\to}A$			δ records location m of first command
1	β ;	$[(0)\overset{+}{\to}A]$			Command enters as $(149)\overset{+}{\to}A$
2	α ;	$(1)\to C$			
3		$p_{11}\overset{\cdot}{\to}C$			Command enters as $(m+1)\to C$
4	α ;	$(C)\to 1$			
5	γ ;	$0\overset{\cdot}{\to}C$		.	Command enters as $100\overset{-}{\to}C$
6		$s(C)\overset{c}{\to}S$			
7	α ;	$1\overset{+}{\to}S$		[1]	Command enters as $m+1\to S$
8		$p_{20}\to T$			

The programme so arranged is listed in Table 3 where the only parameter is the location of the leading command. The command $(149)\overset{+}{\to}A$ is formed from the constant or pseudo-command stored at the foot. The latter is never itself adopted as a command. The command thus formed is substituted in place of the $[p_{20}\to T]$. This is a " stop " command which, if by some mistake not changed, when adopted stops the operation and acts as a warning. The procedure follows as before except that H stores the other variable, 100, and the command $(H_u)\overset{\cdot}{\to}C$ replaces $100\overset{\cdot}{\to}C$.

Routines may be designed in which the roles of parameters and variables are interchangeable. By entering the routines at one or other of two possible points one or other type of data may be adopted. Such routines are used in cases where data may be fixed for one particular programme and then entered as parameters, but may vary during the course of a different calculation and must be set as variables by the programme.

If many variables are required by a routine, beyond the capacity of the arithmetical registers, standard locations in the store may be used, e.g. locations 0, 1, 2, etc. or $m+1$, $m+2$, etc.

VII. The Connection of Routines into Programmes

A complete programme consists primarily of a collection of standard routines selected from a library, together with any routines specially constructed. This part is called the "detail programme". New routines, whenever possible, are constructed in such a manner that they may later be incorporated into the library. The various parts of the detail programme are related to each other by a specially constructed "master programme". The main function of the master programme is to control the course of the calculation and the passage to and from the various routines.

<div align="center">

Table 3

USE OF PROGRAMME VARIABLES

Example from Table 2

</div>

Relative Location of Command		Command as Punched on Tape			Action
0	δ, α	$(10)\overset{+}{\to}A$			Stores head location m and command enters as $(m+10)\to A$ and makes $(A)'=(149)\to A$
1	α	$c(A)\to 2$			Enters as $c(A)\to m+2$, causes $(m+2)'=(149)\to A$, leaving $A'=0$
2		$[p_{20}\to T]$			
3	α	$(2)\to C$			
4		$p_{11}\overset{+}{\to}C$			Has effect of increasing r by unity
5	α	$(C)\to 2$			
6		$(H_u)\overset{-}{\to}C$			Subtracts $100p_{11}$ from C
7		$s(C)\overset{c}{\to}S$			
8	α	$2\to S$	[2]	Return if $r\leqslant 49$
9		$p_{20}\to T$			Stop if $r=50$
10		$<(0)\overset{+}{\to}A>$			Constant or pseudo-command

It is usual for most routines to be separated from the master programme which calls them into use. This is essential for routines which are frequently used or which are used more than once in different contexts by one particular programme.

Some routines, frequently special routines, may be incorporated directly into the master programme, that is, the master programme leads into and out of such routines without special shifts of control. Following EDSAC terminology, such routines are called "open routines".

"Closed routines", into which class the great majority of library routines fall, are entered and left in a special manner involving shifts of control. A convention is adopted, namely, that a closed routine, called into use by one command, finally transfers control back to the command immediately following in the store that from which it was entered. This is not, however, an inflexible rule.

The D register provides storage space for the " linking " data which organize the return of control to the master programme.   Normally the lower-numbered locations in D are reserved as working space used by routines, e.g., D_0, D_1, D_2, and D_3 say.   The higher-numbered locations, D_{11}–D_{15} say, are left free to store linking data.

Thus a routine entered at location t by the command at location m, transfers control back to $m+1$ from, say, location $t+r$.   The set of commands organizing these shifts of control is shown in Table 4.

<div align="center">

TABLE 4

STANDARD LINK PROCEDURE

</div>

Location	Programme Transferring to Routine	Location	Routine
$m-1$	$(S) \to D_{15}$ $\quad\quad\quad \longrightarrow t$	$p_{11} \xrightarrow{+} D_{15}$
m	$t \to S \quad\quad\quad\quad\quad$	—
$m+1$	next command $\longleftarrow\quad t+r$	$(D_{15}) \to S$

The command in $m-1$, known as the " plant " command, places mp_{11}, the content of the sequence register at the instant of transfer, into D_{15} ;   the next command, the " cue ", transfers control to the routine.   The command in t, the " adjustment " command, changes the value of (D_{15}) to $\overline{m+1}\ p_{11}$ whilst the return to $m+1$ is controlled by the " link ", which is the command in $t+r$ by which (D_{15}) is substituted into the sequence register.   As written, the adjustment is the leading command, but this may be placed anywhere convenient within the routine before the link.

Four commands only are needed of which only one, the cue, requires a special datum, the location to which to shift.   All others are standard.   An advantage of this device for linking to and fro is that no arithmetical register normally used for storing variables is required to perform the shifting so that all variables may be set before the cue.

The scheme adopted is very flexible, for control may easily be returned to one of a set of alternative locations.   For instance, before $m-1$ is reached D_{15} may receive a special datum, np_{11} say, which may depend upon the previous course of the calculation.   The command in $m-1$ may be replaced by $(S) \xrightarrow{+} D_{15}$ so that control is returned by the routine to $m+n+1$, where n may take one of a number of values.   Further, routines may sometimes be entered at one of a number of positions and return of control may depend upon the position of entry or upon a variable.   In such cases the adjustment command is replaced by a command which sets the required value in D_{15}.

Some routines may call upon other or auxiliary routines during their operation.   The frequent use of standard routines as auxiliaries to other standard routines helps to reduce the size of a library of routines, since the auxiliary routines may also be used individually.   The method of linking in such cases

may be extended to the auxiliary routines without special reference by the master programme. Those routines which do not use auxiliary routines use D_{15} for linking. Such routines are said to be of the " first " or " lowest order ". Those which use only one additional routine at any instant are of the " second order ". Second order routines may use more than one auxiliary routine but always return control back to the main routine before entering the next auxiliary routine.

Linking into a second order routine is made from the master programme by the use of D_{14}, and such a routine places successive links into D_{15} upon entering each auxiliary routine. " Third order " routines are linked via D_{13} to the master programme and use auxiliary routines linked via D_{14}, which in turn use further routines of first order linked via D_{15}. This procedure may be extended and once the master programme transfers to the highest order routine no reference to the master routine is made during the operation of the routine or its succession

TABLE 5

CONTRACTED LINK PROCEDURE

Location of Command	Command	Action
m	$(S) \rightarrow D_{15}$	Plants $(m+1)p_{11}$ into D_{15}
$m+1$	$p \rightarrow S$	Cue to 1st routine
$m+2$	$q \rightarrow S$	,, ,, 2nd ,,
$m+3$	$r \rightarrow S$	,, ,, 3rd ,,

of auxiliary routines. A routine of lower order may be placed in a standard position relative to the routine using it, e.g. at the end of the latter routine, or the linking data may be supplied to it as a parameter.

Some programmes require the use of a sequence of routines and sub-routines in immediate succession. For a group of routines of the same order, as frequently occurs with first order routines, the linking data may be planted once only for the group by the master routine before entry into the first of the group. Thus, for instance, if a group of first order routines is entered at positions p, q, and r the master programme will contain the set of commands shown in Table 5.

On return of control to $m+2$ the link datum is already stored in D_{15} and is changed to $m+3$ by the adjustment in the routine entered at q and so is already set, upon return to $m+3$, for immediate entry into the third routine at r. This allows of considerable simplification of a master programme.

VIII. FUNCTIONS REQUIRED DURING INPUT OF DATA

During the input of routines some of the data entered must be adjusted to make the routine suited to the programme of which it is a part. The special functions required can be stated.

If a standard routine is available in a medium suitable for machine use, e.g. punched tape, it should be incorporated into the programme as it stands. It

should thus have in it such characters punched as will allow parameters to be introduced. These characters are called " control designations ". Parameters must be supplied to the machine prior to the entry of the routine which requires them and may be stored in groups of standard locations in the store, from which they will be extracted one by one according to the designations detected.

Normally commands will enter sequential store locations. At the commencement of a group of parameters the normal manner of assembly and storage of data must be broken off and the parameters entered into their appropriate positions, after which the normal process must be restored, fresh data being set into the store in locations immediately following the last routine inserted or at any chosen place.

It must also be possible to store the number of the location into which the data being assembled would normally be inserted, and to treat this number as a parameter.

At some stage in the input of a programme it may be necessary to break off the input procedure and to start all or part of the programme so far inserted. This is best done by adopting the last datum assembled from tape as a command.

IX. The Primary and Control Routines

The assembly and storage of data not associated with special designations are performed under the control of the " primary routine ". This is a set of 20 commands which is inserted whenever the computer is started from the " all cleared " condition via a group of stepping switches, where they are permanently wired. With the primary routine stored in locations 0–19 inclusive and the sequence register cleared to zero, data are transferred from punched tape in groups of 10 binary digits per row. A hole is registered as unity and absence of a hole as zero.

As described in Part I each row of the tape possesses 12 positions, 10 binary digits and two controlling positions called X and Y. An X punch is transferred to the input register as a unit p_{19} and a Y as a unit p_{20}.

The action of the primary routine, which is listed in Appendix I, is to add together successive rows which are not X-punched, the partial sums being placed at each stage in the p_{11}–p_{20} group of register A. Upon reading a row possessing an X punch the accompanying 10 digits are added into the p_1–p_{10} group of digits in register A from where it is placed into store.

Before placing a datum into store the next row of holes is " read " and if it possesses no Y punch (A) is then transferred to the store location indicated by the sub-address of the command in location 6. This command is commonly called the " transfer command ". Upon transfer to store the sub-address of (6) is increased by unity by commands 7–9. If, however, a Y is encountered, control is shifted from 3 to 20, the first place beyond the primary routine.

Simple programmes and those which involve only sub-routines can frequently be constructed without the use of parameters, so that the assembly and storage is straightforward. The first command entered is placed in location 20 and this may be the first command of the programme or one causing a shift of control,

and a Y punch, placed on the last row read, will cause control to be shifted to 20 and thus start the programme.

Where parameters are required the first routine entered is the " control routine ". This is assembled in the standard manner and possesses no Y punches and is an open routine. It is entered whenever a Y-punched row is encountered by the primary routine. It is also listed in Appendix I and consists of only 12 commands which are always stored in the locations 20–31 inclusive. Each Y-punched row is accompanied by a group of holes which determines the action taken with regard to the particular parameter already entered or to follow.

Control will be transferred to 20 on the detection of a Y punch. The first command of the control routine clears D_0 of any p_{20} digit it may contain. The next command, 21, stores the sub-address of the transfer command in H. At this stage the p_{20} due to the Y has been removed by the command in 1 and the accompanying digits are *counted* into the sequence register by the command in 22. One of the following actions will take place :

(a) If no units are associated with the Y the control designation is zero ; the next command *adds* (31) to (A) and then shifts control to 9. Then (A) is placed in location 6, which therefore becomes a new transfer command. The sign of D_0 is now zero and control is passed to zero via 11 and so to read the next row of holes. Suppose (A), at the stage of command 13, is np_{11}. Then the addition of (31) causes (A) to become $c(A) \to n$. Eventually the transfer command becomes $c(A) \to n$ and the next data will enter locations n, $n+1$, etc. The number np_{11} may be entered into A from a row preceding the Y-punched row. (A) must be zero before reading n as it will be if the previous command had been normally compiled and stored.

(b) If the Y is accompanied by a single unit, then 23 is omitted and 24 causes the transfer command to be replaced by (A), a datum which may have been compiled from preceding tape rows. It should be noted that the transfer command may be changed by this designation from one causing a storage operation to any other operation, in particular a shift of control. Thus, for instance, the tape command $75 \to S$ followed by this designation will cause (6) to become $75 \to S$. This command will be performed if the next two rows are X-punched. These X punches are needed because the primary routine always reads one column " beyond " the data assembled in order to detect the possible presence of any control designation which may follow and affect the datum just assembled. The same applies to case (a).

(c) When the Y punch is associated with two or more units, one of a different set of operations occurs. In the case of two units control is transferred to 25. This command adds the pseudo-command $<(0) \to T>$ to (A) followed by the addition of $(31) \overset{+}{-} A$; (30) and (23) added together form the command $(H_u) \to 31$. If previous rows have formed np_{11} in A and control moves to 25, then 27 fills 28 with the command $(H_u) \to 31+n$, which is immediately obeyed, and control passes to 29, and thence to read the next row of holes. If n is replaced by a pseudo-command 28 will imply a different operation, e.g. a command demanding a transfer of control.

The example quoted, in which (28) becomes $(H_u) \rightarrow 31 + n$, causes the sub-address of the transfer command to be placed into location $31 + n$. This corresponds to a store location n places beyond the pseudo-command in 31. Locations immediately following 31 are those normally used for storing references of this type which are usually the leading locations of routines.

(d) A Y punch associated with three units causes control to pass to 26, omitting 25. This causes (28) to become $(31 + n) \overset{\pm}{\rightarrow} A$, where (A) is np_{11} at the detection of this control designation. Register A is cleared by command 27 and thus the next row is read with (A) equal to $(31 + n)$. By this means the data stored in a group following a designation of the type (a) or (c) may be selected at will from the store. Thus a command following punched as $25 \rightarrow S$, if preceded by np_{11} and a designation of this type, will be accumulated in A as $\overline{25 + m} \rightarrow S$, where mp_{11} is set by a designation of type (c).

Designations of this kind may be used also to refer to parameters set after a designation of type (a).

(e) When a Y-punched row contains four units control is shifted to 27. Then (A), whatever at this stage it may be, is obeyed as a command since (A) is transferred to 28 by 27. A command may thus be compiled in A by the primary routine and may be obeyed directly upon detection of a designation of this type. In particular this function is useful for producing immediate transfers of control without returning to read further rows, and for rearranging sets of parameters, a function sometimes necessary when the same parameters are needed by more than one routine.

(f) A Y and five units transfers control immediately to 28. This causes the sequence to stop if (28) has been unchanged. The command last placed in 28 by a designation of the types (c), (d), or (e) may be repeated by subsequent designations of this type.

(g) A Y and six units transfers control back to zero and recommences reading with no further action.

(h) A Y and seven units transfers control to 30 and therefore stops the sequence.

(i) A Y and eight units causes control to pass to 31 and thus replaces the first command by whatever is in A.

The cases of nine and ten units would shift control outside the control routine. However, if extra functions are required a special routine may be entered following 31. Allowance for this must be made in placing parameters. Designations of types (g), (h), and (i) are rarely used.

The control routine discussed here is not unique and may be replaced by any other control routine specially constructed. It is usual to adopt that illustrated since designations in standard routines are adopted for this control routine.

X. Use of the Control Routine

The system of control designations resembles the " control combinations " used for programming the EDSAC* and are used in a similar manner. Certain

---

* The two systems were developed independently.

designations are used more frequently than others and designations may follow each other on the tape without the interpolation of commands between them.

Programmes are written in the form in which they are to be punched onto tape, with special symbols representing control designations. Designations are written in the column to the left if preceding, or in the column to the right if following, a datum to which they refer. Sequences of control designations are written in lines as convenient. They are punched in the order they are read from the programme sheet.

The symbols used are as follows :

(a) mT : The symbol m may be replaced by a group of symbols such as a pseudo-command if necessary, but usually will be an integer less than 1024. Thus mT implies changing (6) to $c(A) \rightarrow m$.

(b) R : This symbol is placed after a command or pseudo-command and implies the replacement of (6) by the datum preceding it.

(c) nS : This designation causes the sub-address of the transfer command to be placed in $31+n$ where $n \geqslant 1$.

(d) nA : This causes the addition into register A of $(31+n)$. It usually precedes the datum which it eventually affects. The number n may be replaced by a pseudo-command which may, when added to the command $(31) \stackrel{\rightarrow}{+} A$, form a different command to be obeyed in location 28.

(e) D : This is a single symbol following the datum to which it refers and implies that datum to be obeyed as a command immediately.

(f) This is denoted by U and follows a datum or control designation. It causes repetition of the last nS, nA, or D function performed.

In all these designations the component m or n is punched onto tape as a binary number without an X punch if to be used as a sub-address, or with an X if as a pseudo-command, and immediately precedes the designation proper which is represented by the letter symbol. The letter symbol corresponds to the Y punch and the set of control digits corresponding to the operation required. Thus T is punched as a single hole in the Y position only, R as a Y and one unit and so on. The same keyboard is used for punching both commands and designations.

As an example of the manner in which control designations are used consider the routine listed in Table 2. Clearly the parameters $149p_{11}$ and $100p_{11}$ must precede the routine on the tape. Location 32 is normally reserved for the storage of the head location of the routine entering and 33 onward reserved for other parameters and data such as the locations of heads of other routines. The programme corresponding to Table 2, written in a form suitable for transfer to tape, is shown in Table 6.

Here $m-1$ is the location last filled before the entry of the routine. The first four rows on the tape store m and change (6) so that the two parameters following are entered into locations 33 and 34. These have been written in numerical form and not as pseudo-commands. Then follow two control designations which place mp_{11} in A and change (6) back to $c(A) \rightarrow m$ so that the next command is placed immediately following the last routine, that is, in m.

The routine itself follows on the tape and is now in a standard form. The initial $1S$, which is always placed at the head of each self-referring routine, again places mp_{11} in 32. This value is added into all commands immediately following $1A$ designations. The parameters in 33 and 34 are entered into the commands immediately following the $2A$ and $3A$ designations respectively.

At the foot of the routine the stop command shown in Table 2 is replaced by adjustment and link commands in the standard manner.

TABLE 6

USE OF CONTROL DESIGNATIONS

Location Filled		Tape Entry			Action
		$1S$, $33T$			Store m in 32, changes (6) to $c(A) \rightarrow 33$
33		$149p_{11}$			Parameters
34		$100p_{11}$			
		$1A$, $0T$			Places mp_{11} in A and changes (6) back to $c(A) \rightarrow m$
$m+0$	$1S$	$(A) \overset{-}{\rightarrow} A$			Routine is entered here by programme, mp_{11} sent to 32 again
$m+1$	$2A$	$(0) \overset{+}{\rightarrow} A$			Adds (33) to command and enters as $(149) \overset{+}{\rightarrow} A$
$m+2$	$1A$	$(1) \rightarrow C$			Adds (32) to command and enters as $(m+1) \rightarrow C$
$m+3$		$p_{11} \overset{-}{\rightarrow} C$			
$m+4$	$1A$	$(C) \rightarrow 1$			Adds (32) to command and enters as $(C) \rightarrow m+1$
$m+5$	$3A$	$0 \overset{-}{\rightarrow} C$			Adds (34) to command and enters as $100 \overset{-}{\rightarrow} C$
$m+6$		$s(C) \overset{c}{\rightarrow} S$			
$m+7$	$1A$	$1 \rightarrow S$		[1]	Enters as $m+1 \rightarrow S$ but may be replaced by $-7 \overset{+}{\rightarrow} S$
$m+8$		$p_{11} \overset{+}{\rightarrow} D_{15}$			
$m+9$		$(D_{15}) \rightarrow S$			

XI. The Organization of Complete Programmes

Once the main scheme for carrying out a calculation has been decided upon, the necessary sub-routines and routines may be chosen from the library and special routines may be constructed. Any of these may require parameters and control designations.

The manner and order in which the routines are to be entered into the store are then fixed and a list made of the routines, together with the necessary control designations and parameters. This list describes the detail programme. The master programme must then be designed. This will be largely devoted to the

organization of the course of the programme from one routine to another. The master routine will frequently require a number of control designations associated with it, particularly those giving the positions of the routines of the detail programme. It must be entered into the store after the detail programme and will follow the detail programme on the programme tape.

The construction of a complete programme is illustrated by the example which follows. Suppose we have a tape containing a sequence of decimally punched values of a variable x $(0 \leqslant x < 1)$ and that the corresponding values of $\cos^{-1}[\exp(-\sin \pi x/2)]$ are required to be printed for each entry on the tape. The following routines, which are briefly described, would be required from the library for the detail programme :

(1) Decimal Input A : Sub-routine ; closed. Length ; 22 commands. Entry ; position 0. Link ; D_{15}.

This reads a single positive or negative decimally punched fraction from the tape and converts it to binary form leaving the result in register A.

(2) Exponential A : Sub-routine ; closed. Length ; 13 commands. Entry ; position 0. Link ; D_{15}.

This replaces (A) by its negative exponential, i.e. $(A)' = \exp[-(A)]$.

(3) Sincos A : Routine ; closed ; 1st order. Length ; 33 commands and coefficients. Entry ; position 0 for sine, 1 for cosine. Link ; D_{15}. Parameters ; nil.

This replaces (A) by $\sin \pi(A)/2$ or $\cos \pi(A)/2$ according as the entry is made via the first or leading, or the second command respectively. The routine is headed by the control designation $1S$.

(4) Arcos A : Routine ; closed ; 2nd order. Length ; 19 commands. Entry ; position 0. Link ; D_{14}. Parameters ; location of leading command of sincos A. This routine replaces (A) by its inverse cosine, i.e. $(A)' = 2/\pi.\cos^{-1}(A)$, $-1 \leqslant (A)' < +1$. It uses sincos A which is placed immediately ahead of this routine into which it is linked via D_{15}.

(5) Print A : Sub-routine ; closed. Length ; 33 commands. Entry ; position 0. Link ; D_{15}.

This sub-routine converts positive or negative numbers in A considered as integers (i.e. multiples of 2^{19}) into a series of binary-decimal tetrads which it prints as decimal characters, together with a sign, onto paper. No layout commands are included.

The detail programme is now arranged as in Table 7.

The first control designations cause the first routines to enter from location 38 onwards thus leaving 32–37 free for parameters which are locations of the various leading commands of the routines. These are used by the master programme which is now constructed as illustrated in Table 8.

The first two commands of the master programme enter unaffected and demand a line feed and carriage return of the output printer for which the printer code numbers happen to be 29 and 30. Command No. 3 is changed by the $2A$ designation and shifts control to the decimal input sub-routine. In order to

H

retain the number x entered, it is stored in D_2. Control is moved to No. 4, and x is printed out by the shift to the print routine following. The value of x in A is destroyed by the printing operation and is replaced by command No. 7. Two routines are entered by 9 and 10 without further plant commands. The second order link is planted in D_{14} by 11 and the inverse cosine routine is entered by 12. Commands 13–15 form the product of the computed value in A with the content of the 19th position which converts the result from the grade scale to degree scale. The result is printed and control is returned to the head of the master programme, where the next value is read from the tape.

<div align="center">

TABLE 7

DETAIL PROGRAMME

</div>

Routines and Control Designations	Action upon Entry
Control routine 38T, 2S	Enters positions 20–31 Changes (6) to $c(A) \rightarrow 38$, stores 38p_{11} in 33
Decimal input A 3S	Enters positions 38–59 Stores 60p_{11} in 34
Exponential A 4S	Enters positions 60–72 Stores 73p_{11} in 35
Inverse cosine A	Enters positions 73–91
Sincos A	Enters positions 92–124 ; it is headed by a 1S designation
5S	Stores 125p_{11} in 36
Print A	Enters positions 125–157

It will be noticed that the print routine is used twice, the value from the tape and the corresponding computed value of the function being printed alongside each other. Spacing control is made by the first two commands.

Further, the sincos A routine although used twice is called by the master programme only once, the arcos A routine calling it the second time. This second use is facilitated by the use of a 1A function in the arcos routine which has its position previously placed in location 32 by its initial 1S function.

The last entry of the master programme, which is written as it is to be punched onto tape, calls the computer to transfer control to the head of the master programme, and to start the computation, and to commence the reading of the tape containing the values of x in decimal punch code.

This master programme consists of only 20 commands and controls the passage to and from the detail programme of about 120 commands. Most of the operations of the master programme are associated with the sequence register. This is as it should be with a programme intended to supply the highest order of control to a calculation.

XII. The Construction of Programme and Data Tapes

The computer is capable of accepting data only from punched paper tape and of punching results onto similar tape. Punched cards are used as an auxiliary medium. The two media are associated through the use of special editing equipment used for coding programmes.

TABLE 8

MASTER PROGRAMME

Location Entered	Serial No.	Commands and Control Designations			Action
158	0	$6S$	$29 \rightarrow 0_t$		Stores 158 in 37; causes line feed
159	1		$30 \rightarrow 0_t$		Causes carriage return
160	2		$(S) \rightarrow D_{15}$		Link-plant command
161	3	$2A$	$0 \rightarrow S$	[(33)]	Enters as $38 \rightarrow S$; shifts to " dec. input "
162	4		$(A) \rightarrow D_2$		Stores x in D_2
163	5		$(S) \rightarrow D_{15}$		Link-plant command
164	6	$5A$	$0 \rightarrow S$	[(36)]	Enters as $125 \rightarrow S$; shifts to print values of x
165	7		$(D_2) \rightarrow A$		Restores x to A
166	8		$(S) \rightarrow D_{15}$		Link-plant command
167	9	$1A$	$2 \rightarrow S$	[(34)]	Enters as $94 \rightarrow S$; shifts to compute $\sin \pi(A)/2$
168	10	$3A$	$0 \rightarrow S$	[(34)]	Enters as $60 \rightarrow S$; shifts to compute $\exp [-(A)]$
169	11		$(S) \rightarrow D_{14}$		Link-plant command (2nd order)
170	12	$4A$	$0 \rightarrow S$	[(38)]	Enters as $73 \rightarrow S$; shifts to compute $2/\pi. \cos^{-1}(A)$
171	13	$6A$	$18 \rightarrow C$		Places multiplicand from 177 in C
172	14		$c(A) \overset{\times}{\rightarrow} B$		Forms product to reduce
173	15		$(R) \overset{+}{\underset{-}{\rightarrow}} A$		(A) to (degrees of arc) $\times 10^3 \times 2^{-19}$ and rounds off
174	16		$(S) \rightarrow D_{15}$		Link-plant command
175	17	$5A$	$0 \rightarrow S$	[(36)]	Enters as $125 \rightarrow S$; shifts to print result as an integer
176	18	$6A$	$0 \rightarrow S$	[(37)]	Enters as $158 \rightarrow S$; returns control to head of master programmes
177	19		$< 0 \cdot 1716614 >$		Constant used to convert result to degrees of arc, equal to $9 \times 10^{-3} \times 2^{-19}$
		$6A$	$0 \rightarrow S$	D	Enters as $158 \rightarrow S$ and obeyed immediately

The library of routines is stored on punched cards of the usual Hollerith or I.B.M. type. Each routine or sub-routine occupies one or more cards, and commands and binary data are punched in a columnar fashion in groups of 10 digits, the address group being X-punched. Control designations are included. Punched columns on each card are terminated by a single column punched with an X and Y only. This punch, when read by a card reader, causes the card to be ejected and the next in the deck fed into the reading position. The card reader consists of an electric duplicating key punch in which the punches have been replaced by reading brushes.

The advantage of the use of cards is that they are easily stored in a small space and that many copies of each routine or sub-routine may be stored and replenished when required. Damage sustained in use by such copies is of little importance.

After the design of the detail and master programmes has been completed the necessary routine cards are selected from the library and placed into the card reader in the order in which they are to be transferred to tape. Blank tape is fed into the tape punch. The routines are copied from the cards directly onto the tape column for row except for the final XY of each card. Data such as control designations and parameters are supplied via the keyboard between successive routines. Finally, the master programme is punched via the keyboard together with open routines which it may incorporate. The latter are reproduced from cards if standard. Before a tape is used a second tape is prepared in the same way and the two are compared by a special tape comparator. This ceases comparing when any corresponding non-identically punched rows are detected. Punching errors, but not programming errors, can then be eliminated. Tapes may also be duplicated by connecting a tape reader to the editing tape punch in place of the card reader. This also allows correct portions of tapes to be copied and errors to be corrected from the keyboard.

Cards are also used to provide decimal data. This provision is due to the great distances between populated areas in Australia. A distant user may punch decimal data onto cards in the conventional manner. These data are transferred for input directly onto tape without translation of code.

Results may be punched onto tape in the same code as cards are decimally punched. After the output tape has been edited the results are transferred to cards through a tape reader and a special card punch. The cards may then be sorted, reproduced, and listed in any desired manner. This procedure is useful for the production of printed tables. The form of the page may be chosen and the cards may then be listed in a suitable manner for reproduction as a volume. The card punch is used also for recording routines onto cards for the library from the keyboard.

XIII. CHECKING

The computer possesses no independent or automatic checking devices. All responsibility for adequate checking of results is left to the programmer. There are, of course, two main aims in checking, firstly to ensure that the programmes are correctly designed and, secondly, to ensure the absence of

instrumental faults.   Special devices are applicable to one or other of these aims, although in some circumstances the same device may be used for the detection of either type of error.

Much source of error in the construction of programmes is diminished by the use of library routines whose properties are known precisely.   However, although the correct routines may be chosen for a particular calculation, errors in the assembly of data and in the master programme may exist and editing checks must always be applied before placing the tape into the machine.   Coding errors may most easily arise from false keying on the keyboard.   These are detected by comparing a pair of independently punched tapes.   Similarly, before placing a new routine in the library two sets of cards are punched independently and are compared before reproduction in bulk.

Each routine or sub-routine is tested on the machine before being placed into the library.   In such tests registers N_1 and N_2 are useful for setting data to be used by the routine being tested.   The manual control made possible by these registers allows of the rapid detection of logical faults in both routines under test and in full-scale programmes.

Although it is possible to go through each new programme being tested by performing each command one by one (depression of a " single operation " key on the control panel allows one command to be obeyed) and inspecting the register content, such a procedure is very tedious.   Programmes are usually broken into distinct parts during testing, each part being separated by a stop command.   These commands are placed in positions at which the content of the registers may be of particular importance.

The special stop commands may, when a programme is known to be correct, be replaced by " record " commands which transfer the content of one register to another to be used for inspection, e.g. a spare position in register D or the store.   The contents of registers A, B, C, and H are shown on cathode-ray oscillograph faces as also is the entire content of D and any group of 16 consecutive locations of the store.   The state of a calculation may be seen by a glance at the tube faces.

Special routines are sometimes used in tracing logical faults in programmes. These are less highly developed than are checking routines used in the EDSAC mainly because their use has not been found essential in this machine.   This is not because such faults have not occurred but that the more elementary devices of viewing results, using stop and record commands together with the use of the hand-set registers N_1 and N_2, have been adequate.

Of the few checking routines used, one will cause the calculation to proceed through the programme tested and will print out the locations of the programme at which a shift of control occurs.   The route taken through the programme may then be traced.   Another routine will print the content of registers A, B, and C over a chosen range of the programme.

Programmes, once known to be correct in the store, are page printed by use of a special routine.   A record of the entire programme in its final form is obtained in case repunching should be necessary.   The programme prints the addresses in a letter code as close to the written source and destination symbols as possible,

The sub-address is printed in the scale of 32 for simplicity in repunching. A programme of similar type can be used to punch a tape of the final programme, as it lies in the store.   This is particularly useful in cases of programmes required in identical form at a later date.

To check that no electrical faults develop during running of a calculation, programmes are designed to include whatever mathematical checks may be applied to the calculation, and special calculations whose results are known are repeated at intervals.   In calculations which have short programmes applied to large amounts of data, selected samples of the data are submitted to calculation before proceeding to the main bulk of the work.   The results for the selected data must be reproduced in the main calculation.

Before daily use all units of the machine are subjected to standard tests. The early tests are manually controlled.   Later stages, designed to test the arithmetical registers and function gates, involve specially designed programmes.

XIV. Use of the Low Speed Store

The low speed store comprises four independent groups of 1024 locations with an access time of 10 msec.   It is used for storing incidental results, matrices of coefficients, tables of functions, approximations to solutions to partial differential equations, etc.

One of these stores is used to hold one or more programmes which may be in use.   Thus if a programme is to be used for a number of days it is convenient to hold it in one of these stores, preferably in locations corresponding to those in which it lies in the high speed store when in operation.   When required the programme is transferred to the high speed store.   This takes only a few seconds.

Although the access time of this store is 10 msec, use of it during calculation does not seriously reduce the rate of operation.   This is due partly to its infrequent use even if referred to for much arithmetical work.   The speed of operation is controlled primarily by the access time to commands and not to calculation data, to which relatively few commands refer.   This is so since the major part of a programme is concerned with its own control, a function which does not require the low speed store.

Programmes which are too extensive to be held entirely within the high speed store may be adopted for action from the low speed store as required. This may be done in two ways, either by taking the commands one by one from the low speed store and obeying them with the aid of a special routine or by transferring portions of the programme from the low speed store to the high speed store, as and when required for immediate use.   The latter method is that usually adopted, for greater operating speed can be attained if each routine is used for relatively long periods before being over-written by the next routine. The proportion of time spent in transferring routines is then small.

Programmes may be placed in the low speed store directly from the programme tape.   When they are transferred to the high speed store for use the space normally occupied by the primary and control routines and the parameters may be occupied by the programme placed in the low speed store from location zero onward.   This saves up to 40 locations in the high speed store.

When programmes are withdrawn from the low speed store in blocks, each block is assembled from the tape into the high speed store and then transferred to the low speed store by a special routine before proceeding to reading in the next block of commands.

XV. Manual Controls

The manual controls include those for starting and stopping the computer at will, for clearing registers, and for setting certain 20-digit data. Others, intended to help the operator, allow the computer to operate in special ways.

Three rows of 20 switches each provide for the adjustment of the content of registers N_1 and N_2 and of the input register. Register N_1 possesses an added facility which allows its content to be accepted as a command irrespective of the content of the sequence register. This allows the operator to change the content of any arithmetical register or of the stores at will, and so allows of manual adjustment of programmes when corrections to them are needed.

Besides the independent application of individual commands via N_1, a succession of commands differing only by sequential sub-addresses may be accepted. The sub-address in each case is the current content of the sequence register. By use of this facility data from the low speed store may be transferred to the high speed store or vice versa item by item, and the whole of the content of a store may be transferred within a few seconds without the use of any additional programme. This facility is also useful in maintenance tests.

A further device called a " trigger unit ", useful for maintenance tests and in programme tests, provides a cathode-ray oscillograph trigger pulse whenever the content of the sequence register coincides with that set on a special group of switches or will stop the computer so that the content of the register may be inspected.

XVI. References

PEARCEY, T., and HILL, G. W. (1953).—*Aust. J. Phys.* **6** : 316.

WILKES, M. V., WHEELER, D. J., and GILL, S. (1951).—" The Preparation of Programmes for an Electronic Digital Computer." p. 10. (Addison-Wesley Press Inc. : Cambridge, Mass.)

APPENDIX I

PRIMARY ROUTINE AND CONTROL ROUTINES USED DURING THE ENTRY OF DATA INTO THE COMPUTER
Primary routine : locations 0–19 ; control routine : locations 20–31

Position of Command	Command			Action
0	$(I) \rightarrow C$			Reads column of 10 digits, X, Y to register C
1	$p_{2_0} \xrightarrow{+} C$			Alters most significant digit (Y)
2	$p_{20} \cdot (C) \xrightarrow{c} S$			Tests most significant digit
3	$20 \rightarrow S$		[20]	Shifts to 20 if Y on current column
4	$p_{20} \cdot (D_0) \xrightarrow{c} S$			If no Y punch, tests most significant digit of D_0

APPENDIX I (*Continued*)

Position of Command	Command			Action
5	$12 \rightarrow S$	[12]		If no p_{20} in D_0, shifts to 12
6	$[c(A) \rightarrow 20]$			" Transfer command "—plants A and clears A ; called if previous column X-punched and current column no X or Y
7	$(6) \rightarrow A$			⎫ Increases store address of transfer command by unity
8	$p_{11} \overset{+}{\rightarrow} A$			⎬
9	$c(A) \rightarrow 6$			⎭
10	$p_{20}.(D_0) \overset{c}{\rightarrow} S$			Tests most significant digit of D_0 again, p_{20} if previous column X-punched and current column no X.
11	$0 \rightarrow S$	[0]		Returns to read if no X on previous column
12	$p_{19} \overset{-}{\rightarrow} C$			
13	$p_{20}.(C) \rightarrow D_0$			If current column X-punched, p_{20} in C is read to D_0 ; if no X, D_0 cleared
14	$(C) \rightarrow H_l$			⎫ Clears out p_{19} and p_{20} digits from C and leaves 10 digits in C and H
15	$(Hl) \rightarrow C$			⎭
16	$p_{20}.(D_0) \overset{c}{\rightarrow} S$			Tests for X on current column
17	$(H_u) \rightarrow C$			If no X, shift column digits to p_{11}–p_{20} groups
18	$(C) \overset{+}{\rightarrow} A$			If no X, add 10 digits p_{11}–p_{20} to A
19	$0 \rightarrow S$	[0]		Return to " read "
20	$(D_0) \overset{-}{\rightarrow} D_0$			Clears D_0 for fresh assemblage to follow
21	$(6) \rightarrow H_u$			Stores address of transfer command sent to H register
22	$(C) \overset{c}{\rightarrow} S$			Digits associated with Y punch counted to sequence register
23	$(31) \overset{+}{\rightarrow} A$			If $(A) = n$, then $(A)' = c(A) \rightarrow n$
24	$9 \rightarrow S$	[9]		Causes transfer command to be replaced by (A) ; $(6)' = (A)$
25	$(30) \overset{+}{\rightarrow} A$			This and following cause $H_u \rightarrow 31 + (A)$ to be placed in 28
26	$(23) \overset{+}{\rightarrow} A$			Causes $(31) \overset{+}{\rightarrow} A + (A)$ to be placed in 28
27	$c(A) \rightarrow 28$			Assembled (A) planted for use
28	$p_1 \rightarrow T$			Command assembled placed here and performed
29	$0 \rightarrow S$	[0]		Return to continue reading data
30	$<(0) \rightarrow T>$			Constant when added to (23) forms $(H_u) \rightarrow 31$
31	$<c(A) \rightarrow M>$			Constant used by command 23

SHORT COMMUNICATIONS

ON THE LOW EXCITED STATE OF Li⁷ †

By N. W. TANNER‡ and R. G. UEBERGANG‡

From the mass of evidence reflecting on the low excited state of Li⁷, it seems very probable that the spin is either $\frac{1}{2}$ or 5/2. The main experimental result favouring $\frac{1}{2}$ is the isotropy of the γ-radiation, and this from several reactions. Inglis (1951) considered the α-γ correlation of B¹⁰(n, α)Li⁷*γLi⁷ (Rose and Wilson 1950) and concluded that the observed isotropy could only be reconciled with spin 5/2 if the γ-radiation is a particular mixture of magnetic dipole and electric quadrupole.

Subsequent correlation experiments (Newton 1951 ; Class and Hanna 1952 ; Uebergang and Tanner 1953) have all indicated isotropy. The simplest conclusion is spin $\frac{1}{2}$ for the decaying state. If spin 5/2 is assumed for Li⁷* then the experimental results can be satisfied under the conditions :

(a) Li⁷* decays by the same multipole mixture notwithstanding the way in which it is formed.

(b) A given interfering multipole mixture can cause isotropy of the γ-radiation for all ways in which the 478 keV level is excited and for all methods of observing the radiation.

The first condition is assumed to be true *a priori.* The second is suggested by the approximate formulae of Lloyd (1951) and the explicit calculations of Devons (1949) and Ling and Falkoff (1949).

Lloyd has considered the modification of a calculated X-γ correlation (pure magnetic γ-radiation) to the case in which a small admixture of electric radiation occurs ; only interference terms are introduced, the pure electric component being neglected. For the radiation of interest here, magnetic dipole plus electric quadrupole, the condition for isotropy through multipole interference reduces to a simple expression depending only on the parameters of the γ-decay, and not on the way in which the decaying state is formed.

More generally one may consider a state of spin J, with degenerate magnetic substates designated by the quantum numbers m, decaying to a state of spin J'. It is fundamental in angular correlations that the magnetic substates m are not equally populated and that the relative populations $P(m)$ depend on the way in which the state is formed. For simplicity the phases of the substates are assumed to be random ; this is satisfied providing the previous radiations involved in

the correlation are parallel or anti-parallel and are taken as the axis of quantiza-tion (see Biedenharn, Arfken, and Rose 1951). If the γ-decay is an interfering mixture of 2^L-pole magnetic and 2^{L+1}-pole electric radiation, then according to Ling and Falkoff (1949) the angular distribution of that radiation is given by

$$W(\theta) = \sum_{Mm} [P(m)\{|\alpha|^2(JLmM \mid J'M+m)^2 F_L^M(\theta)$$
$$+|\beta|^2(JL+1mM \mid J'M+m)^2 F_{L+1}^M(\theta)$$
$$+(\alpha\beta^*+\alpha^*\beta)(JLmM \mid J'M+m)(JL+1mM \mid J'M+m)F_{L,L+1}^M(\theta)\}],$$
$$\cdots\cdots\cdots\cdots\cdots\cdots (1)$$

where M is the magnetic quantum number associated with the γ-radiation,

       α, β are the complex amplitudes of the dipole and quadrupole radiations respectively,

     $F_L^M(\theta)$ is the classical angular distribution of 2^L-pole γ-radiation (magnetic or electric),

$F_{L,L+1}^M(\theta)$ is the classical angular distribution of the interference component.

The remaining factors are transformation coefficients for vector addition as defined and tabulated by Condon and Shortley (1935).

For the case of the α-γ correlation of $B^{10}(n, \alpha)Li^{7*}\gamma Li^7$, Devons (1949) had shown that a quadrupole intensity admixture of about 4 per cent. can cause isotropy. It seems then, that terms in $|\beta|^2$ can be omitted. Making this necessary simplification, the angular distribution of radiation $W_m(\theta)$ from one of the degenerate magnetic substates m is given by

$$W_m(\theta) = \sum_M \{(JLmM \mid J'M+m)^2 F_L^M(\theta)$$
$$+2\delta(JLmM \mid J'M+m)(JL+1mM \mid J'M+m)F_{L,L+1}^M(\theta)\},$$
$$\cdots\cdots\cdots\cdots\cdots\cdots (2)$$

where

$$2\delta = \frac{\alpha\beta^*+\alpha^*\beta}{|\alpha|^2}.$$

Evaluation of equation (2) for the case of $L=1$ and $J'=J-1$ leads to the isotropy condition

$$\delta = \tfrac{1}{2}\left(\frac{2J+2}{30(J-1)}\right)^{\frac{1}{2}}, \quad \cdots\cdots\cdots\cdots (3)$$

which is independent of m. Hence isotropy through multipole interference does not depend on the populations of the substates, that is, the way in which the decaying state is formed or observed.

Applying (3) to Li^{7*} with $J=5/2$ gives $\delta=1/5$ in approximate agreement with the formulae of Devons (1949) and also Lloyd (1951). It remains then that, although a 4 per cent. relative quadrupole intensity is considered unlikely (Inglis 1951), an unexpected effect may be occurring which causes isotropy of the 478 keV γ-radiation by multipole interference. Assignment of spin $\tfrac{1}{2}$ to the state in question does not seem to be completely beyond doubt.

References

BIEDENHARN, L. C., ARFKEN, G. B., and ROSE, M. E. (1951).—*Phys. Rev.* **83** : 586.

CLASS, C. M., and HANNA, S. S. (1952).—*Phys. Rev.* **87** : 247.

CONDON, E. U., and SHORTLEY, G. H. (1935).—" The Theory of Atomic Spectra." (Cambridge Univ. Press.)

DEVONS, S. (1949).—*Proc. Phys. Soc. Lond.* A **62** : 580.

INGLIS, D. R. (1951).—*Phys. Rev.* **81** : 914.

LING, D. S., and FALKOFF, D. L. (1949).—*Phys. Rev.* **76** : 1639.

LLOYD, S. P. (1951).—*Phys. Rev.* **83** : 716.

NEWTON, J. O. (1951).—*Proc. Phys. Soc. Lond.* A **64** : 938.

ROSE, B., and WILSON, A. R. W. (1950).—*Phys. Rev.* **78** : 68.

UEBERGANG, R. G., and TANNER, N. W. (1953).—*Aust. J. Phys.* **6** : 53.

EFFICIENCIES IN THE METHOD OF GROUPING

By P. G. Guest*

Summary

The efficiencies obtained in curve fitting by the method of grouping are discussed in terms of two parameters $\varkappa_2$, $\varkappa_3$ which specify the departure from uniform spacing. For polynomials of the first and second degree the efficiencies practically always exceed $0 \cdot 7$, but the efficiencies for the third degree polynomial may be less than this value if the spacing is markedly non-uniform.

I. Introduction

In an earlier paper (Guest 1952) a method of fitting polynomials to unequally spaced observations was described, the method being named the method of grouping. Although intended for use in cases where the spacing was non-uniform, it was only possible at that time to discuss the efficiencies for cases in which the variation from uniformity was random and the standard errors did not differ markedly from the errors in the equally spaced case.

Since the publication of this paper a method of treating cases in which the spacing is non-uniform has been devised, the departure from uniform spacing being characterized by two parameters $\varkappa_2$, $\varkappa_3$. The behaviour of the standard errors in the least squares problem has been described in terms of the two parameters (Guest 1953). In the present paper the calculation of the efficiencies of the values obtained by the method of grouping will be carried out in terms of these same parameters.

II. Calculation of the Standard Errors

The coefficients b_{pj} in the fitted polynomial

$$u_p(x) = \sum_{j=0}^{p} b_{pj} x^j$$

are determined in the method of grouping by the solution of the " normal " equations

$$\sum_r W_k(x) \left\{ y(x) - \sum_{j=0}^{p} b_{pj} x^j \right\} = 0, \qquad k = 0 \text{ to } p, \quad \cdots\cdots\cdots \quad (1)$$

where the $y(x)$ represent the observations and the $W_k(x)$ are step functions. The methods of solving these equations consist in eliminating in turn the coefficients b_{p0}, b_{p1}, etc. The most convenient method is some variant of what Dwyer

---

* Physics Department, University of Sydney.

A

(1951) calls the method of single division.   Equations (1) are in effect converted to the set

$$\sum_x W_{k.k}(x)\left\{y(x)-\sum_j b_{pj}x^j\right\}=0, \qquad k=0 \text{ to } p, \quad \ldots\ldots (2)$$

where the functions $W_{k.k}(x)$ are linear combinations of the functions $W_k(x)$ such that

$$\sum_x W_{k.k}(x)x^m=0, \quad m<k. \ldots\ldots\ldots\ldots\ldots\ldots (3)$$

$W_{k.k}(x)$ may be expanded in the form

$$W_{k.k}(x)=W_k(x)+\sum_{m=0}^{k-1}\alpha_{km}W_{m.m}(x), \quad \ldots\ldots\ldots (4)$$

and it follows from (3) that

$$\alpha_{km}=-\sum W_k(x)x^m/\sum W_{m.m}(x)x^m. \quad \ldots\ldots\ldots (5)$$

The value α_{kk} is defined to be -1.   From the coefficients α_{km} the coefficients β_{km} defined by the equations

$$W_{k.k}(x)=\sum_{m=0}^k \beta_{km}W_m(x) \quad \ldots\ldots\ldots\ldots\ldots (6)$$

may be derived.   In fact

$$\beta_{km}=\sum_{r=m}^{k-1}\alpha_{kr}\beta_{rm}. \quad \ldots\ldots\ldots\ldots\ldots (7)$$

For the coefficient of degree p, equation (2) with $k=p$ gives

$$b_{pp}=\sum_x W_{p.p}(x)y(x)\bigg/\sum_x W_{p.p}(x)x^p,$$

and so

$$\sigma^2(b_{pp})/\sigma^2(y)=\sum_x\left[W_{p.p}(x)\right]^2\bigg/\left[\sum_x W_{p.p}(x)x^p\right]^2. \quad \ldots\ldots\ldots (8)$$

The expression $\sum_x\left[W_{p.p}(x)\right]^2$ can be evaluated from (6) when the values of $\sum_x W_k{}^2(x)$, $\sum_x W_k(x)W_m(x)$ are known.

If the efficiencies of the other coefficients b_{pj} or the efficiencies of the fitted values are required, it is necessary to complete the inversion of the matrix $\sum W_k(x)x^m$.   Functions $W_{k.p}(x)$ may be defined for which

$$\sum_x W_{k.p}(x)x^m=0, \qquad m\leqslant p, \qquad m\neq k, \quad \ldots\ldots (9)$$

and then

$$b_{pk}=\frac{\sum W_{k.p}(x)y(x)}{\sum W_{k.p}(x)x^k}=\sum_x\sum_{m=0}^p \lambda_{km}W_m(x)y(x), \quad \ldots\ldots (10)$$

where λ_{km} are the elements of the inverse matrix. These elements can be built up from the quantities β_{jk}, $\Sigma W_{j.j}(x)x^k$ in the following way. $W_{k.p}(x)$ is expanded in the form

$$W_{k.p}(x) = W_{k.k}(x) + \sum_{k+1}^{p} \alpha_{km} W_{m.p}(x),$$

where, from (9)

$$\alpha_{km} = -\frac{\Sigma W_{k.k}(x)x^m}{\Sigma W_{m.p}(x)x^m} = -\frac{\Sigma W_{k.k}(x)x^m}{\Sigma W_{m.m}(x)x^m}.$$

Therefore

$$W_{k.p}(x)\Big/\sum_{x} W_{k.p}(x)x^k = \left[W_{k.k}(x) - \sum_{m}\left\{\sum_{x} W_{k.k}(x)x^m\right\}\frac{W_{m.p}(x)}{\sum_{x} W_{m.m}(x)x^m}\right]\Big/\sum_{x} W_{k.k}(x)x^k$$

$$= \left[\sum_{r}\beta_{kr}W_r(x) - \sum_{m}\left\{\sum_{x} W_{k.k}(x)x^m\right\}\sum_{r}\lambda_{mr}W_r(x)\right]\Big/\sum_{x} W_{k.k}(x)x^k$$

$$= \sum_{r}\lambda_{kr}W_r(x).$$

Hence

$$\lambda_{kr} = \left[\beta_{kr} - \sum_{k+1}^{p}\lambda_{mr}\left\{\sum_{x} W_{k.k}(x)x^m\right\}\right]\Big/\sum_{x} W_{k.k}(x)x^k. \quad \ldots\ldots\ldots \text{ (11)}$$

The standard error of the coefficient b_{pk} is given by

$$\sigma^2(b_{pk})/\sigma^2(y) = \sum_{x}\left\{\sum_{m=0}^{p}\lambda_{km}W_m(x)\right\}^2, \quad \ldots\ldots\ldots\ldots \text{ (12)}$$

and the standard error of the fitted value by

$$\sigma^2[u_p(x)] = E\left[\sum_{j} b_{pj}x^j\right]^2 = \sum_{j, k} x^j x^k E[b_{pj}b_{pk}],$$

or

$$\sigma^2[u_p(x)]/\sigma^2(y) = \sum_{j, k} x^j x^k \sum_{x}\left\{\sum_{r}\lambda_{jr}W_r(x)\right\}\left\{\sum_{s}\lambda_{ks}W_s(x)\right\}. \quad \ldots\ldots \text{ (13)}$$

The method of matrix inversion outlined above is the same as that described by Fox and Hayes (1951), but they deal with a general matrix for which the functions $W_j(x)$ are not defined. The present discussion brings out the significance of the intermediate terms occurring in their method. In the usual method of calculation the quantities are arranged in a square array, as shown below:

$$
\begin{array}{llll}
\Sigma W_{0.0} & \Sigma W_{0.0}x & \Sigma W_{0.0}x^2 & \Sigma W_{0.0}x^3 \\
(-)\alpha_{10} & \Sigma W_{1.1}x & \Sigma W_{1.1}x^2 & \Sigma W_{1.1}x^3 \\
(-)\alpha_{20} & (-)\alpha_{21} & \Sigma W_{2.2}x^2 & \Sigma W_{2.2}x^3 \\
(-)\alpha_{30} & (-)\alpha_{31} & (-)\alpha_{32} & \Sigma W_{3.3}x^3.
\end{array}
$$

The lower triangular matrix $(-)\underset{\sim}{\alpha}$ is then inverted; the elements of $-\underset{\sim}{\alpha}^{-1}$ are the quantities β_{km}, as is clear from (7). Finally the rows λ_{jk} are built up in turn, beginning with λ_{pk}. However, when a large number of inversions have to be made it is more convenient to tabulate the α_{km}, β_{km}, etc., in columns and perform the same calculations for all the matrices at the same time.

It is illuminating to put some of the above equations into matrix notation. The quantities $\sum\limits_{x} W_j(x)x^k$ form a matrix $\mathbf{W}$, the quantities α_{jk} a lower triangular matrix $\underset{\sim}{\alpha}$, and the quantities $\sum\limits_{x} W_{j,j}(x)x^k$ an upper triangular matrix $\underset{\sim}{\omega}$. Then, from equation (4)

$$\sum_{x} W_{k,k}(x)x^r = \sum_{x} W_k(x)x^r + \sum_{m=0}^{k-1}\alpha_{km}\sum_{x} W_{m,m}(x)x^r,$$

or

$$\mathbf{W} = -\underset{\sim}{\alpha}\underset{\sim}{\omega}.$$

Then from equation (7)

$$\underset{\sim}{\alpha}\underset{\sim}{\beta} = -\mathbf{I},$$

and so

$$\underset{\sim}{\beta} = -\underset{\sim}{\alpha}^{-1}.$$

Finally equation (11) may be written

$$\omega_{kk}\lambda_{kr} = \beta_{kr} - \sum_{m}\omega_{km}\lambda_{mr},$$

and so

$$\underset{\sim}{\beta} = \underset{\sim}{\omega}\underset{\sim}{\lambda},$$

and

$$\underset{\sim}{\lambda} = \underset{\sim}{\omega}^{-1}\underset{\sim}{\beta} = -\underset{\sim}{\omega}^{-1}\underset{\sim}{\alpha}^{-1}$$
$$= \mathbf{W}^{-1}.$$

III. Efficiencies in Terms of the Parameters $\varkappa_2$, $\varkappa_3$

The symbol ε will be used to denote the variable which takes the integral or half-integral values from $+\frac{1}{2}(n-1)$ to $-\frac{1}{2}(n-1)$ at the points of observation, where n is the number of observations. In the method of grouping (Guest 1952, Section III)

$$\sum_{\varepsilon} W_j \varepsilon^m = \frac{n^{m+1}}{(m+1)2^m}[1 - \alpha_j^{m+1} - \beta_j^{m+1} + \dots\dots], \quad m+j \text{ even}, \\ = 0, \hspace{5.5cm} m+j \text{ odd}, \left.\right\} \dots (14)$$

where α_j, β_j, etc., are the parameters which determine the location of the groups. Equation (14) can be written in the form

$$\sum_{\varepsilon} W_j \varepsilon^m = \frac{n^{m+1}}{(m+1)2^m}\psi_{jm}.$$

The quantities ψ_{jm} may be readily calculated from the values of the parameters α_j, β_j, etc., given in the previous paper (Guest 1952). The numerical values of ψ_{jm} are listed in Table 1. If the variable $e = 2\varepsilon/n$ is introduced, then

$$n^{-1}\Sigma W_j e^m = \psi_{jm}/(m+1). \quad \dots\dots\dots\dots (15)$$

In accordance with the treatment given in an earlier paper (Guest 1953, p. 132), the independent variable x is replaced by the variable ξ, where

$$\xi = \varepsilon + \varkappa_2 n^{-1}\left(\varepsilon^2 - \frac{1}{12}n^2\right) + 2\varkappa_3 n^{-2}\left(\varepsilon^3 - \frac{1}{4}n^2\varepsilon\right), \quad \dots\dots (16)$$

and $\varkappa_2$, $\varkappa_3$ are the parameters which specify the departure from uniform spacing.

Now

$$12n^{-1}\xi = 6e + \varkappa_2(3e^2 - 1) + 3\varkappa_3(e^3 - e),$$

and so

$$n^{-1}\Sigma W_j(12n^{-1}\xi) = 3\psi_{j1} + \varkappa_3\left(\frac{3}{4}\psi_{j3} - \frac{3}{2}\psi_{j1}\right) + \varkappa_2(\psi_{j2} - \psi_{j0})$$

$$n^{-1}\Sigma W_j(12n^{-1}\xi)^2 = 12\psi_{j2} + \varkappa_3\left(\frac{36}{5}\psi_{j4} - 12\psi_{j2}\right) + \varkappa_3{}^2\left(\frac{9}{7}\psi_{j6} - \frac{18}{5}\psi_{j4} + 3\psi_{j2}\right)$$

$$+ \varkappa_2{}^2\left(\frac{9}{5}\psi_{j4} - 2\psi_{j2} + \psi_{j0}\right)$$

$$+ \varkappa_2[(9\psi_{j3} - 6\psi_{j1}) + \varkappa_3(3\psi_{j5} - 6\psi_{j3} + 3\psi_{j1})],$$

$$n^{-1}\Sigma W_j(12n^{-1}\xi)^3 = 54\psi_{j3} + \varkappa_3(54\psi_{j5} - 81\psi_{j3}) + \varkappa_3{}^2\left(\frac{81}{4}\psi_{j7} - 54\psi_{j5} + \frac{81}{2}\psi_{j3}\right)$$

$$+ \varkappa_3{}^3\left(\frac{27}{10}\psi_{j9} - \frac{81}{8}\psi_{j7} + \frac{27}{2}\psi_{j5} - \frac{27}{4}\psi_{j3}\right)$$

$$+ \varkappa_2{}^2\left[(27\psi_{j5} - 27\psi_{j3} + 9\psi_{j1}) + \varkappa_3\left(\frac{81}{8}\psi_{j7} - \frac{45}{2}\psi_{j5} + \frac{63}{4}\psi_{j3} - \frac{9}{2}\psi_{j1}\right)\right]$$

$$+ \varkappa_2\left[\left(\frac{324}{5}\psi_{j4} - 36\psi_{j2}\right) + \varkappa_3\left(\frac{324}{7}\psi_{j6} - \frac{432}{5}\psi_{j4} + 36\psi_{j2}\right)\right.$$

$$+ \varkappa_3{}^2(9\psi_{j8} - 27\psi_{j6} + 27\psi_{j4} - 9\psi_{j2})$$

$$\left. + \varkappa_2{}^2\left(\frac{27}{7}\psi_{j6} - \frac{27}{5}\psi_{j4} + 3\psi_{j2} - \psi_{j0}\right)\right].$$

Using Table 1, it is a simple matter to calculate $\Sigma W_j\xi^m$ as a function of $\varkappa_2$, $\varkappa_3$. These expressions are tabulated in Table 2, together with the quantities $\Sigma W_j{}^2$, $\Sigma W_j W_k$, which are needed for the evaluation of the standard errors.

TABLE 1

THE QUANTITIES ψ_{jm}

				$p=1, 2$	$p=3$		
ψ_{20}	—0·2943	ψ_{11}	+0·888889	+0·502270	ψ_{31}	—0·514459
ψ_{22}	+0·385331	ψ_{13}	+0·987654	+0·752265	ψ_{33}	—0·181496
ψ_{24}	+0·669354	ψ_{15}	+0·998628	+0·876695	ψ_{35}	+0·078761
ψ_{26}	+0·809080	ψ_{17}	+0·999848	+0·938627	ψ_{37}	+0·275442
ψ_{28}	+0·885873	ψ_{19}	+0·999983	+0·969453	ψ_{39}	+0·425384

TABLE 2

$\sum W_j \xi^m$

$n^{-1}\sum W_0$	1
$n^{-1}\sum W_1$	0
$n^{-1}\sum W_2$	—0·2943
$n^{-1}\sum W_3$	0
$4n^{-2}\sum W_0\xi$	0
$4n^{-2}\sum W_1\xi$	$0\cdot888889 - 0\cdot197531\varkappa_3 \quad (p=1, 2)$
	$0\cdot502270 - 0\cdot063069\varkappa_3 \quad (p=3)$
$4n^{-2}\sum W_2\xi$	$0\cdot226544\varkappa_2$
$4n^{-2}\sum W_3\xi$	$-0\cdot514459 + 0\cdot211856\varkappa_3$
$12n^{-3}\sum W_0\xi^2$	$1 - 0\cdot400000\varkappa_3 + 0\cdot057143\varkappa_3{}^2 + 0\cdot066667\varkappa_2{}^2$
$12n^{-3}\sum W_1\xi^2$	$\varkappa_2[0\cdot296296 - 0\cdot021948\varkappa_3] \quad (p=1, 2)$
	$\varkappa_2[0\cdot313064 - 0\cdot031391\varkappa_3] \quad (p=3)$
$12n^{-3}\sum W_2\xi^2$	$0\cdot385331 + 0\cdot016281\varkappa_3 - 0\cdot017786\varkappa_3{}^2 + 0\cdot011656\varkappa_2{}^2$
$12n^{-3}\sum W_3\xi^2$	$\varkappa_2[0\cdot121108 - 0\cdot018177\varkappa_3]$
$32n^{-4}\sum W_0\xi^3$	$\varkappa_2[0\cdot533333 - 0\cdot076190\varkappa_3 + 0\cdot008466\varkappa_2{}^2]$
$32n^{-4}\sum W_1\xi^3$	$0\cdot752265 - 0\cdot251702\varkappa_3 + 0\cdot039489\varkappa_3{}^2 - 0\cdot002379\varkappa_3{}^3$
	$\qquad\qquad + \varkappa_2{}^2[0\cdot145927 - 0\cdot011742\varkappa_3]$
$32n^{-4}\sum W_2\xi^3$	$\varkappa_2[0\cdot546337 - 0\cdot120582\varkappa_3 + 0\cdot013561\varkappa_3{}^2 + 0\cdot017713\varkappa_2{}^2]$
$32n^{-4}\sum W_3\xi^3$	$-0\cdot181496 + 0\cdot351005\varkappa_3 - 0\cdot111592\varkappa_3{}^2 + 0\cdot012001\varkappa_3{}^3$
	$\qquad\qquad + \varkappa_2{}^2[0\cdot044385 + 0\cdot008764\varkappa_3]$
$n^{-1}\sum W_0{}^2$	1
$n^{-1}\sum W_1{}^2$	$0\cdot666667 \quad (p=1, 2)$
	$0\cdot2945 \quad (p=3)$
$n^{-1}\sum W_2{}^2$	$0\cdot7265$
$n^{-1}\sum W_3{}^2$	$0\cdot7894$
$n^{-1}\sum W_0 W_2$	—0·2943
$n^{-1}\sum W_1 W_3$	—0·0417

The standard errors of the coefficients b_{pp} can be calculated for selected values of $\varkappa_2$, $\varkappa_3$ from the values $\sum W_j \xi^k$, using the scheme developed in Section II. The efficiencies can then be calculated by comparison with the corresponding errors obtained for the least squares curve in the earlier paper. The efficiencies obtained in this way for the coefficients b_{11}, b_{22}, b_{33}, are listed in Table 3.

In Table 4 the efficiencies of the fitted values have been tabulated for various values of $\varkappa_2$, $\varkappa_3$. For the second and third degree polynomials the variable used is

$$k = e - \varkappa_2/5,$$

which was introduced in the treatment of the least squares standard errors (Guest 1953, equation (**41**)). The range of interpolation, that is, the range of values of k within which the observations occur, is roughly from $k = +1$ to $k = -1$.

TABLE 3

EFFICIENCIES OF THE COEFFICIENTS b_{pp}

$\varkappa_3$ / $\varkappa_2^2$	b_{11}					b_{22}					b_{33}		
	0	0·5	1·0	1·5	2·0	0	0·25	0·5	0·75	1·0	0	0·25	0·5
−1·0	0·911	0·891	0·871	0·853	0·835	0·853	0·801	0·751	0·705	0·660	0·813	0·641	0·498
−0·8	0·909	0·887	0·866	0·847	0·828	0·872	0·819	0·768	0·720	0·675	0·861	0·692	0·548
−0·6	0·906	0·882	0·860	0·839	0·819	0·887	0·834	0·783	0·734	0·687	0·897	0·735	0·593
−0·4	0·901	0·876	0·853	0·830	0·809	0·897	0·844	0·793	0·744	0·696	0·921	0·769	0·632
−0·2	0·896	0·869	0·844	0·820	0·798	0·902	0·851	0·801	0·751	0·703	0·934	0·793	0·663
0	0·889	0·860	0·833	0·808	0·784	0·902	0·853	0·804	0·755	0·707	0·937	0·807	0·685
0·2	0·880	0·849	0·821	0·794	0·769	0·897	0·851	0·803	0·755	0·708	0·929	0·810	0·699
0·4	0·869	0·836	0·806	0·777	0·751	0·888	0·844	0·799	0·752	0·705	0·913	0·804	0·702
0·6	0·855	0·820	0·788	0·758	0·731	0·874	0·833	0·790	0·745	0·698	0·888	0·789	0·696
0·8	0·839	0·801	0·767	0·736	0·707	0·855	0·818	0·777	0·733	0·688	0·856	0·765	0·681
1·0	0·818	0·779	0·743	0·710	0·680	0·833	0·799	0·760	0·718	0·673	0·817	0·733	0·657
1·2	0·794	0·752	0·715	0·681		0·808	0·776	0·739	0·698	0·653	0·770	0·693	0·626
1·4	0·764	0·721	0·682			0·780	0·749	0·713	0·672	0·628	0·715	0·645	0·587
1·6	0·729	0·684	0·644			0·751	0·719	0·683	0·641	0·596			
1·8	0·688	0·642	0·602			0·719	0·686	0·647	0·604	0·556			
2·0	0·640	0·594	0·554			0·687	0·649	0·606	0·558	0·507			

In a practical example the efficiencies may be expected to differ somewhat from the values given in Tables 3 and 4, because of the neglect in the present discussion of higher parameters $\varkappa_4$, $\varkappa_5$, etc. In Table 5 are shown the efficiencies of the coefficients b_{pp} for the three examples discussed in an earlier paper (Guest 1953), with the values calculated from the parameters $\varkappa_2$, $\varkappa_3$ in brackets. It is seen that there is in each case a reasonable agreement between the two values for the efficiency.

IV. CONCLUSION

From Table 3 it will be seen that the efficiencies of the coefficients b_{pp} are always less than the corresponding efficiencies in the equally spaced case ($\varkappa_2 = 0$, $\varkappa_3 = 0$), except for the coefficient b_{11} when $\varkappa_3$ is negative, where the efficiency may be slightly higher. As a consequence, the value of 90 per cent. suggested in the earlier paper for the efficiencies must be regarded as an upper limit to the efficiencies which would be found in any practical example.

TABLE 4
EFFICIENCIES OF THE FITTED VALUES
First Degree Polynomial

$|x_2|$	0					0·5					1·0				
x_3 $|e|$	−1·0	−0·5	0	+0·5	+1·0	−1·0	−0·5	0	+0·5	+1·0	−1·0	−0·5	0	+0·5	+1·0
1·4	0·923	0·917	0·903	0·880	0·840	0·914	0·906	0·891	0·865	0·822	0·888	0·875	0·854	0·821	0·772
1·2	0·927	0·920	0·908	0·885	0·847	0·918	0·910	0·895	0·871	0·830	0·893	0·881	0·860	0·829	0·781
1·0	0·932	0·926	0·914	0·893	0·857	0·924	0·916	0·903	0·879	0·840	0·900	0·889	0·870	0·840	0·794
0·8	0·940	0·935	0·924	0·905	0·873	0·933	0·926	0·914	0·893	0·857	0·912	0·901	0·884	0·857	0·815
0·6	0·952	0·948	0·939	0·924	0·897	0·946	0·941	0·931	0·913	0·884	0·929	0·920	0·906	0·883	0·848
0·4	0·969	0·967	0·961	0·951	0·933	0·966	0·962	0·955	0·944	0·924	0·954	0·949	0·939	0·924	0·899
0·2	0·990	0·989	0·987	0·983	0·977	0·988	0·987	0·985	0·981	0·974	0·984	0·982	0·979	0·973	0·964
0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Second Degree Polynomial

$|x_2|$	0					0·5					1·0				
x_3	−1·0	−0·5	0	+0·5	+1·0	−1·0	−0·5	0	+0·5	+1·0	−1·0	−0·5	0	+0·5	+1·0
$|k|$						(kx_2 negative)									
1·4	0·868	0·898	0·903	0·880	0·833	0·835	0·871	0·887	0·877	0·838	0·726	0·757	0·775	0·775	0·748
1·2	0·879	0·902	0·904	0·881	0·835	0·854	0·883	0·896	0·885	0·846	0·766	0·786	0·799	0·796	0·765
1·0	0·901	0·911	0·908	0·884	0·838	0·893	0·905	0·911	0·898	0·858	0·855	0·847	0·844	0·831	0·795
0·8	0·940	0·930	0·918	0·893	0·848	0·950	0·942	0·935	0·919	0·879	0·962	0·948	0·927	0·899	0·852
0·6	0·952	0·952	0·939	0·915	0·874	0·938	0·951	0·950	0·940	0·911	0·872	0·914	0·943	0·957	0·942
0·4	0·924	0·946	0·950	0·938	0·912	0·886	0·914	0·926	0·927	0·921	0·779	0·815	0·840	0·859	0·877
0·2	0·904	0·935	0·946	0·941	0·925	0·863	0·896	0·911	0·911	0·904	0·750	0·785	0·807	0·816	0·810
						(kx_2 positive)									
0	0·898	0·931	0·943	0·939	0·924	0·861	0·898	0·916	0·918	0·906	0·754	0·793	0·821	0·834	0·826
0·2	0·904	0·935	0·946	0·941	0·925	0·876	0·913	0·934	0·936	0·920	0·786	0·831	0·865	0·883	0·876
0·4	0·924	0·946	0·950	0·938	0·912	0·911	0·941	0·951	0·942	0·910	0·858	0·902	0·929	0·937	0·918
0·6	0·952	0·952	0·939	0·915	0·874	0·950	0·948	0·928	0·895	0·848	0·942	0·946	0·927	0·894	0·848
0·8	0·940	0·930	0·918	0·893	0·848	0·922	0·902	0·877	0·842	0·794	0·894	0·860	0·819	0·775	0·725
1·0	0·901	0·911	0·908	0·884	0·838	0·865	0·862	0·848	0·819	0·774	0·786	0·769	0·743	0·708	0·663
1·2	0·879	0·902	0·904	0·881	0·835	0·833	0·845	0·839	0·813	0·768	0·726	0·725	0·710	0·683	0·641
1·4	0·868	0·898	0·903	0·880	0·833	0·818	0·838	0·837	0·813	0·768	0·697	0·705	0·697	0·675	0·635

Third Degree Polynomial

$|x_2|$	0					0·5				
x_3	−1·0	−0·5	0	+0·5	+1·0	−1·0	−0·5	0	+0·5	+1·0
$|k|$			(kx_2 negative)							
1·4	0·830	0·910	0·926	0·887	0·805	0·625	0·747	0·818	0·826	0·779
1·2	0·842	0·909	0·919	0·881	0·802	0·643	0·756	0·823	0·832	0·787
1·0	0·870	0·902	0·905	0·871	0·802	0·736	0·790	0·836	0·842	0·802
0·8	0·874	0·889	0·887	0·870	0·834	0·872	0·897	0·881	0·866	0·834
0·6	0·844	0·902	0·917	0·914	0·901	0·711	0·835	0·895	0·903	0·884
0·4	0·852	0·914	0·936	0·932	0·908	0·695	0·816	0·891	0·918	0·911
0·2	0·879	0·925	0·942	0·936	0·913	0·759	0·858	0·919	0·942	0·939
			(kx_2 positive)							
0	0·898	0·931	0·943	0·939	0·924	0·889	0·927	0·938	0·923	0·886
0·2	0·879	0·925	0·942	0·936	0·913	0·907	0·915	0·903	0·873	0·824
0·4	0·852	0·914	0·936	0·932	0·908	0·821	0·860	0·871	0·864	0·836
0·6	0·844	0·902	0·917	0·914	0·901	0·767	0·834	0·872	0·891	0·894
0·8	0·874	0·889	0·887	0·870	0·834	0·820	0·874	0·888	0·875	0·843
1·0	0·870	0·902	0·905	0·871	0·802	0·887	0·879	0·850	0·795	0·714
1·2	0·842	0·909	0·919	0·881	0·802	0·800	0·836	0·826	0·774	0·688
1·4	0·830	0·910	0·926	0·887	0·805	0·752	0·813	0·817	0·772	0·687

TABLE 5

EFFICIENCIES IN PRACTICAL EXAMPLES

Values calculated from $\varkappa_2$, $\varkappa_3$ in brackets

Example	n	$\varkappa_2^2$	$\varkappa_3$	b_{11}	b_{22}	b_{33}
1	16	0·015	—0·642	0·878 (0·906)	0·878 (0·881)	0·902 (0·879)
2	67	0·345	+0·740	0·825 (0·818)	0·824 (0·806)	0·751 (0·739)
3	16	0·221	—0·392	0·884 (0·890)	0·839 (0·850)	0·823 (0·788)

The effect of departures from uniform spacing may be roughly summarized in the following way :

Departure from Uniformity		Efficiency		
		b_{11}	b_{22}	b_{33}
Slight	$\lvert \varkappa_2 \rvert$, $\lvert \varkappa_3 \rvert < 0·25$	>0·875	>0·875	>0·900
Moderate	$\lvert \varkappa_2 \rvert$, $\lvert \varkappa_3 \rvert < 0·5$	>0·850	>0·840	>0·750
Pronounced	$\lvert \varkappa_2 \rvert$, $\lvert \varkappa_3 \rvert < 0·75$	>0·800	>0·750	>0·550

Since the efficiency of the fitted value $u_b(x)$ is at worst only slightly less than the efficiency of the coefficient b_{pp}, the limiting efficiencies of the fitted values will also be given roughly by the above table.   However, from Table 4 it will be seen that the efficiency of the fitted value varies quite rapidly with the location of the point (i.e. the coordinate k) in the second and third degree polynomials when the departure from uniformity becomes pronounced.

The value that would be considered acceptable for the efficiency depends very much on the purpose for which the curve is required.   If the curve is to summarize the results of 6 months' research, then clearly the least squares curve should be calculated.   If a large number of curves are to be plotted, then the method of grouping may well be more appropriate because of the saving in time. Jeffreys' (1948) statement on this point is worth quoting in full.

" If [the estimate] a' has an efficiency of 50 per cent., a' will habitually differ from [the least squares estimate] a by more than the standard error of the latter.   This is very liable to be serious.   No general rule can be given ; we have in particular cases to balance accuracy against the time that would be needed for an accurate calculation, but as a rough guide it may be said that efficiencies over 90 per cent. are practically always acceptable, those between 70 and 90 per cent. usually acceptable, but those under 50 per cent. should be avoided."

Cases in which $\lvert \varkappa_2 \rvert$ or $\lvert \varkappa_3 \rvert$ exceed unity will be very rare.   It can be said then that the efficiencies for polynomials of the first and second degree fall into the " usually acceptable " category, while for the third degree polynomial the efficiencies will only fall into this category when the departure from uniformity is not very pronounced.

For the fourth degree polynomial the representation in terms of the two parameters $\varkappa_2$, $\varkappa_3$ is not very satisfactory, but a calculation of the standard errors

for the case $\varkappa_3 = 0$ has shown that the drop in efficiency as $|\varkappa_2|$ increases is even more pronounced than is the case with the third degree polynomial.   Consequently the method of grouping should not be used with a polynomial of the fourth degree unless the spacing is roughly uniform.

V. REFERENCES

DWYER, P. S. (1951).—" Linear Computations."   (John Wiley & Sons : New York.)

Fox, L., and HAYES, J. G. (1951).—*J. R. Statist. Soc.* B **13** : 83.

GUEST, P. G. (1952).—*Aust. J. Sci. Res.* A **5** : 238.

GUEST, P. G. (1953).—*Aust. J. Phys.* **6** : 131.

JEFFREYS, H. (1948).—" Theory of Probability."   2nd Ed.   p. 179.   (Oxford Univ. Press.)

INTERNAL CONVERSION IN THE L–SUBSHELLS

By J. B. Swan* and R. D. Hill†

Summary

The relative intensities of conversion-electron lines from the three L-subshells have been determined for a number of γ-transitions, and the results compared with the theoretical values of Gellman, Griffith, and Stanley (1952). Experiment has been shown to be in accord with theory.

Calculations of theoretical K/L ratios indicate that the empirical curves of Goldhaber and Sunyar (1951) may need to be modified, especially for low Z. For a given Z^2/E, the K/L ratio increases with decreasing atomic number.

I. INTRODUCTION

Accurate theoretical values of the internal conversion coefficients of electric and magnetic multipole radiation for the L-subshells would be of great assistance in the identification of transitions, and in particular of those which, owing to their low energy, do not convert in the K-shell.

Conversion coefficients for the K-shell have been computed for a wide range of energies, Z-values, and multipolarities by Rose, Goertzel, and Perry (1951) and Rose *et al.* (1951) in the relativistic case with the unscreened Coulomb field acting on the electron, and for a more restricted range, but including screening, by Reitz (1950).

Calculations of L-shell conversion coefficients have been made by Hebb and Nelson (1940), Tralli and Lowen (1949), and more recently by Gellman, Griffith, and Stanley (1952). The results of Gellman, Griffith, and Stanley are of particular interest, as they have been computed for the three L-subshells for E1, E2, and M1 radiations, using relativistic wave functions for the atomic electrons but not taking into account the effects of screening. As these results are the most complete at present available, it is of interest to determine the degree of reliability which may be attached to them.

An experimental investigation of the relative L-subshell conversions has been made by Mihelich (1952), whose results in specific cases support those of Gellman, Griffith, and Stanley (1952). It should be pointed out that the latter's calculations indicate certain crossings of the L-subshell conversion curves at particular values of Z and E. In the case of E2 transitions, for example, at high Z the L_I conversion can be more important than the L_{II} conversion. This

* Fulbright Fellow ; present address : Physics Department, University of Western Australia, Nedlands, W.A.

† Physics Department, University of Illinois, Urbana, Ill.

was already pointed out by Mihelich in a note added in proof after Gellman, Griffith, and Stanley's paper had appeared, but nevertheless the generalization for all electric transitions implied by Mihelich in the abstract of his paper has often since been misinterpreted.

In the present paper, results of investigations of several additional transitions are presented, and these are compared with Gellman, Griffith, and Stanley's values, both for relative $L_I : L_{II} : L_{III}$ ratios and for absolute conversion coefficients. In order that a valid comparison may be made, it is desirable that the nature of the transition involved be known beforehand. Several transitions have been selected, for which the L-subshell conversion lines can be clearly resolved, and the identification of the transition has been made independently on the basis of lifetime, K-conversion, and K/L ratio.

II. IDENTIFICATION OF TRANSITION MULTIPOLARITIES
$^{181}_{73}$Ta : 133 keV

This transition arises in the $2 \cdot 2 \times 10^{-5}$ sec isomeric state of ^{181}Ta following the β-decay of ^{181}Hf, and has been quoted by Goldhaber and Hill (1952) as E2. The theoretical radiation lifetimes τ_γ, calculated by Weisskopf (1951) and expressed in nomograph form by Montalbetti (1952), are $1 \cdot 4 \times 10^{-11}$ sec for M1 and $1 \cdot 4 \times 10^{-7}$ sec for E2. The experimental value, calculated from $\tau_\gamma = T_{\frac{1}{2}}(1 + \alpha_{\text{total}})/\ln 2$ by employing a measured total conversion coefficient of ~ 1 (Hedgran and Thulin 1951 ; Chang-Yun Fan 1952), is $\sim 6 \cdot 5 \times 10^{-5}$ sec, which, although long, is in reasonable agreement with an E2 assignment, and seems to rule out an appreciable M1, E2 mixture. The theoretical K-conversion coefficients of Rose, Goertzel, and Perry (1951) and Rose et al. (1951) are $1 \cdot 91$ for M1, $0 \cdot 50$ for E2. Experimental values of $0 \cdot 34$ (Chang-Yun Fan 1952) and $0 \cdot 49$ (Hedgran and Thulin 1951) indicate a pure E2 transition. The experimental K/(L+M) ratio obtained by both these authors is $0 \cdot 61$, in excellent agreement with the empirical ratio of Goldhaber and Sunyar (1951) of $0 \cdot 6$ for E2.

The absence of any M1 component is unexpected in view of the spin assignments quoted by Goldhaber and Hill (1952), but the transition is probably to be identified as pure E2.

$^{186}_{76}$Os : 137 keV

This transition arises in the 8×10^{-10} sec isomeric state of ^{186}Os following the β-decay of ^{186}Re, and has been quoted by Goldhaber and Hill (1952) as E2. The theoretical radiation lifetimes (Weisskopf 1951 ; Montalbetti 1952) are $1 \cdot 3 \times 10^{-11}$ sec for M1 and 9×10^{-8} sec for E2. The experimental value, employing a measured total conversion coefficient of ~ 1 (Metzger and Hill 1951 ; Steffen 1951), is $\sim 2 \cdot 3 \times 10^{-9}$ sec, which is in satisfactory agreement with pure E2, though it may indicate some admixture of M1. The theoretical K-conversion coefficients of Rose, Goertzel, and Perry (1951) and Rose et al. (1951) are $2 \cdot 17$ for M1, $0 \cdot 43$ for E2. Experimental values of $0 \cdot 35$ (Metzger and Hill 1951) and $0 \cdot 37$ (Steffen 1951) indicate an E2 transition with no M1

admixture. The empirical K/L ratios cannot be used for identification in this case, since the E2 curve has been drawn utilizing this transition. The observed K/L ratio is $0 \cdot 6$ (Metzger and Hill 1951 ; Steffen 1951). It seems reasonable to identify this transition as pure E2.

$^{191}_{77}$Ir : 129 keV

This transition follows the β-decay of ^{191}Os, and has been quoted by Goldhaber and Hill (1952) as an M1+E2 mixture. No lifetime has been measured. The theoretical radiation lifetimes (Weisskopf 1951 ; Montalbetti 1952) are $1 \cdot 7 \times 10^{-11}$ sec for M1 and $1 \cdot 0 \times 10^{-7}$ sec for E2. The theoretical K-conversion coefficients of Rose, Goertzel, and Perry (1951) and Rose et al. (1951) are $3 \cdot 0$ for M1, $0 \cdot 49$ for E2. The experimental value of $1 \cdot 36$ (Kondaiah 1951 ; Swan and Hill 1952) indicates a $0 \cdot 42 : 0 \cdot 58$ mixture of M1 : E2 transitions. The empirical K/L ratios of Goldhaber and Sunyar (1951) are $7 \cdot 5$ for M1 and $0 \cdot 36$ for E2. The observed value of $2 \cdot 2$ (Kondaiah 1951 ; Swan and Hill 1952) indicates a $0 \cdot 64 : 0 \cdot 36$ mixture of M1 : E2. While the evidence does not agree accurately on the relative M1 : E2 mixture, it is clear that this transition does involve a mixing of M1 and E2 radiations.

$^{198}_{80}$Hg : 411 keV

This transition follows the β-decay of ^{198}Au, and has been identified by Hill and Mihelich (1950) as E2. No lifetime has been measured, though the observations of Graham and Bell (1951), Moon (1951), and Bell, Graham, and Petch (1952) indicate a half-life of $\sim 10^{-11}$ sec. The theoretical radiation lifetimes (Weisskopf 1951 ; Montalbetti 1952) are 5×10^{-13} sec for M1 and 4×10^{-10} sec for E2. The experimental value, employing a total conversion coefficient of $\sim 0 \cdot 1$, is $\sim 2 \times 10^{-11}$ sec, which is consistent with pure E2. The theoretical K-conversion coefficients of Rose, Goertzel, and Perry (1951) and Rose et al. (1951) are $0 \cdot 17$ for M1, $0 \cdot 032$ for E2. The observed values of $0 \cdot 04$ (Plesset 1942), $0 \cdot 025$ (Peacock and Wilkinson 1948), and $0 \cdot 03$ (Siegbahn and Hedgran 1949) indicate a pure E2 transition. The empirical K/L ratios cannot be used in identification, for the reasons given under ^{186}Os. The observed value is $2 \cdot 1$ (Hill and Mihelich 1950). The transition may be identified as probably pure E2.

$^{199}_{80}$Hg : 159 keV

This transition arises in the $2 \cdot 4 \times 10^{-9}$ sec isomeric state of ^{199}Hg, following the β-decay of ^{199}Au, and has been quoted by Goldhaber and Hill (1952) as E2. The theoretical radiation lifetimes (Weisskopf 1951 ; Montalbetti 1952) are 1×10^{-11} sec for M1 and 5×10^{-8} sec for E2. The experimental value, employing a measured total conversion coefficient of $0 \cdot 6$ (Sherk and Hill 1951), is $5 \cdot 5 \times 10^{-9}$ sec, which is in satisfactory agreement with pure E2, though it may indicate some admixture of M1 radiation. The theoretical K-conversion coefficients of Rose, Goertzel, and Perry (1951) and Rose et al. (1951) are $2 \cdot 24$ for M1, $0 \cdot 28$ for E2. The experimental value of $0 \cdot 19$ (Sherk and Hill 1951) indicates that no M1

is present.   The empirical K/L ratios cannot be used in identification, for the reasons given under ^{186}Os.   The observed values are 0·8 (Beach, Peacock, and Wilkinson 1949), 0·87 (Hill and Mihelich 1950), and 0·6 (Sherk and Hill 1951). The transition may be identified as probably pure E2.

$$^{203}_{81}\text{Tl} : 279 \text{ keV}$$

This transition follows the β-decay of ^{203}Hg, and has been identified as E2 (Saxon 1948 ; Wilson and Curran 1951).   No lifetime has been measured, though Deutsch and Wright (1950) have shown the half-life to be shorter than 3×10^{-9} sec.   The theoretical radiation lifetimes (Weisskopf 1951 ; Montalbetti 1952) are 2×10^{-12} sec for M1 and 3×10^{-9} sec for E2.   The observed upper limit to the half-life is too long to allow of any differentiation between M1 and E2.   The theoretical K-conversion coefficients of Rose, Goertzel, and Perry (1951) and Rose et al. (1951) are 0·535 for M1, 0·079 for E2.   The experimental values of 0·18 (Saxon 1948 ; Wilson and Curran 1951) and 0·15 (Thulin, personal communication from I. Bergström, 1953), using a mean of 0·16, indicate a 0·24 : 0·76 mixture of M1 : E2.   The empirical K/L ratios of Goldhaber and Sunyar (1951) are 7·7 for M1 and 1·3 for E2.   The observed values of 3 (Saxon 1948 ; Slätis and Siegbahn 1949a, 1949b), 3·7 (Wilson and Curran 1951) and 3·5 (Thulin, personal communication from I. Bergström, 1953), using a mean of 3·5, indicate a 0·40 : 0·60 mixture of M1 : E2, a lower K/L ratio giving a smaller percentage of M1.   It is clear that this transition does involve a mixing of M1 and E2 radiations.

III. Experimental Data

The relative L-subshell conversion ratios of ^{181}Ta, ^{186}Os, ^{191}Ir, and ^{203}Tl were measured using a photographic 180° magnetic spectrograph of 0·1 to 0·2 per cent. resolution, and are given in Table 1.   The values tabulated for ^{198}Hg and ^{199}Hg are those of Hill and Mihelich (1950), originally assigned to conversion in the L_I- and L_{III}-shells, but later by Mihelich (1952) to L_{II} and L_{III}.

The figures for the theoretical L-subshell conversion were obtained from Gellman, Griffith, and Stanley's (1952) table by interpolation and extrapolation. For each multipolarity and value of the atomic number Z a log-log plot of the conversion coefficient v. γ-energy, k, in units of mc^2, was prepared.   The conversion coefficients so obtained for the particular k were plotted against Z on a log-linear scale, and interpolation made for the particular Z.

The theoretical K/L ratio is the ratio of Rose's K-conversion coefficient to Gellman, Griffith, and Stanley's total L-conversion coefficient.   Since it can be assumed that the true K-conversion is given accurately by Rose's K-conversion coefficients (Rose, Goertzel, and Perry 1951 ; Rose et al. 1951), in comparing this theoretical ratio with either the observed or empirical K/L ratio, one is in effect comparing theoretical and observed absolute total L-conversion.   Alternatively, an absolute L-conversion coefficient may be obtained from K-conversion and K/L ratio data, and values have been tabulated in Table 2 for the E2 transitions.   In this table " experimental " values have also been calculated

TABLE 1

THEORETICAL AND EXPERIMENTAL L-CONVERSION COEFFICIENTS AND K/L RATIOS

Converting Nucleus	Transition Energy (keV)	Theoretical L-Subshell Conversion						Observed Ratio			Theoretical K/L Ratio		Empirical K/L Ratio		Observed K/L Ratio
		M1			E2						M1	E2	M1	E2	
		L_I	L_{II}	L_{III}	L_I	L_{II}	L_{III}	L_I	L_{II}	L_{III}					
^{181}Ta	133	0·34 340 :	0·018 18 :	0·001 1·0	0·064 0·21 :	0·36 1·2 :	0·30 1·0	0·20 :	1·22 :	1·00	5·3	0·69	7·5	0·6	0·61
^{186}Os	137	0·42 420 :	0·025 25 :	0·001 1·0	0·058 0·17 :	0·40 1·14 :	0·35 1·0	small :	1·24 :	1·00	4·9	0·54	7·5	0·6	0·6
^{198}Hg	411	0·028 280 :	0·0018 18 :	0·0001 1·0	0·0060 2·73 :	0·0060 2·73 :	0·0022 1·0	— :	2·5* :	1·00	5·7	2·3	7·8	2·6	2·1
^{199}Hg	159	0·405 405 :	0·027 27 :	0·001 1·0	0·042 0·192 :	0·294 1·34 :	0·219 1·0	small :	1·6 :	1·00	5·7	0·51	7·5	0·6	0·6
^{191}Ir	129	0·52 370 :	0·033 24 :	0·0014 1·0	0·069 0·147 :	0·58 1·23 :	0·47 1·0	4·9 :	1·75 :	1·00	5·4	0·44	7·5	0·36	2·2
^{203}Tl	279	0·086 308 :	0·0075 27 :	0·00028 1·0	0·013 0·81 :	0·030 1·88 :	0·016 1·0	2·85 :	2·08 :	1·00	5·7	1·34	7·7	1·3	3·5

* There appears to be only the L_{II} line present. However, the L_I and L_{II} lines would barely be resolved on account of the relatively high γ-energy.

using the theoretical K-conversion coefficient as, in general, K/L ratios may be determined with greater accuracy than K-conversion coefficients, particularly if the decay scheme is complex.

TABLE 2

L-CONVERSION COEFFICIENTS OF E2 TRANSITIONS

Converting Nucleus	L-Conversion Coefficient		
	Theoretical	$\dfrac{\alpha_{K \text{ expt.}}}{(K/L)_{\text{expt.}}}$	$\dfrac{\alpha_{K \text{ th.}}}{(K/L)_{\text{expt.}}}$
^{181}Ta	0·72	0·80	0·82
^{186}Os	0·81	0·60	0·73
^{198}Hg	0·014	0·014	0·015
^{199}Hg	0·56	0·32	0·47

IV. DISCUSSION
E2

The values given in Tables 1 and 2 indicate that for the pure E2 radiations investigated Gellman, Griffith, and Stanley's (1952) values of L-conversion coefficients are in accord with experiment. The observed relative L_{II}-conversion of ^{186}Os and ^{199}Hg may be in even better agreement with theory if one considers that the L_{II}-intensity has been increased by a small contribution from the weak L_I-conversion line. However, the case of ^{198}Hg is more difficult to explain. Here the measured L_{II}/L_{III} ratio is 2·5, in agreement with theory, but no L_I-line has been observed. This line should, according to theory, be of intensity comparable with that of the L_{II}-line. However, with the present resolution of $\sim$0·2 per cent., a broader line somewhere between the L_I and L_{II} positions might be expected with approximately the observed intensity. Regarding the absolute L-conversion coefficient, in ^{186}Os and ^{199}Hg, where the K-conversion is lower than the theoretical value, it is apparent from Table 2 that this has the effect of similarly reducing the calculated experimental L-conversion coefficient.

M1+E2

Due to the uncertainty in the relative M1 and E2 transition intensities of M1+E2 mixtures, it is more profitable to calculate the relative intensities from the observed and theoretical L-conversion ratios ; by comparison with mixtures determined by alternative means, the reliability of the theoretical L-conversion coefficients may be decided. For the 129-keV transition in ^{191}Ir, the observed L_I/L_{III} ratio and Gellman, Griffith, and Stanley's values indicate a 0·8 : 0·2 mixture of M1 : E2, which gives an L_{II}/L_{III} ratio of $\sim$1·4, in fair accord with the experimental ratio of 1·75. The agreement with either a 0·42 : 0·58 or a 0·64 : 0·36 mixture, by K-conversion and K/L ratios respectively, is not satisfactory, but merely serves to confirm existence of the mixture. It is unlikely that the poor agreement results from the neglect of (M+N)-conversion in the calculation of total transition intensity. For the 279-keV transition in ^{203}Tl,

the observed L_I/L_{III} ratio indicates a $0 \cdot 35 : 0 \cdot 65$ mixture of M1 : E2, giving an L_{II}/L_{III} ratio of $2 \cdot 04$, in excellent agreement with the experimental ratio of $2 \cdot 08$. Satisfactory agreement also exists with mixtures determined by K-conversion and K/L ratios.

In calculating mixtures, it should be pointed out that we have assumed strictly " linear " mixtures of the mixing multipoles, proportional to the total transition probabilities, that is, including K- and L-electron conversion transitions as well as unconverted γ-ray transitions. The effect on the mixtures of including (M+N)-conversion transitions is slight, amounting to only 1 or 2 per cent., as is the effect of using the K/L ratios discussed in Section V, in place of the empirical values.

V. Theoretical K/L Ratios

Since it appears from Table 2 that there are reasonable grounds for considering that the absolute magnitudes of the total L-shell conversions given by Gellman, Griffith, and Stanley (1952) are approximately correct, it would be valuable for experimental purposes to have plots of the K/L ratio for M1 and E2 transitions. It will be recalled that Gellman, Griffith, and Stanley computed
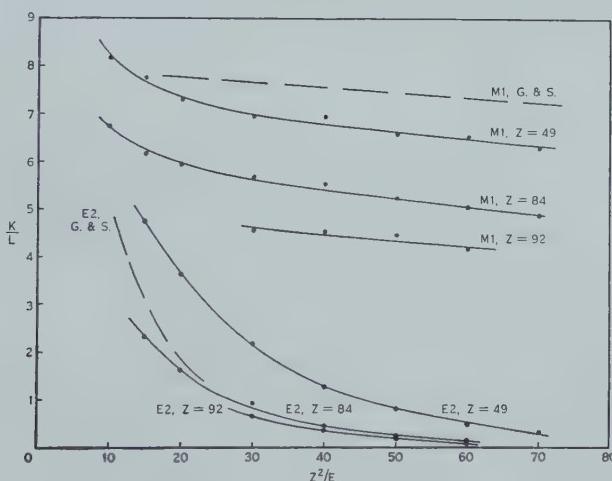
Fig. 1.—Theoretical K/L ratios for M1 and E2 transitions for different values of Z. The curves of Goldhaber and Sunyar (1951), G. & S., are included for comparison. These curves may also be compared with those of Hebb and Nelson (1940), and Tralli and Lowen (1949), calculated for Z=35.

the L-conversion coefficients for the same Z-values (49, 84, and 92) as used by Reitz (1950) in the latter's computation of K-conversion coefficients. The K/L ratios plotted in Figure 1 have been calculated from the Reitz and Gellman, Griffith, and Stanley values and, following the usual practice, have been plotted v. Z^2/E (E in keV).

It will be noticed that, while the K/L ratio curves show the accepted trends of M1 and E2 transition, there is a significant difference between the curves of different Z-values. In view of the fact that Goldhaber and Sunyar (1951) gave

B

only one curve for each multipole transition, it is at first sight strange that the behaviour suggested by the curves of Figure 1 has not been seen earlier. These authors did point out that it would be very likely that the exact K/L ratios would depend on Z and E in a complicated manner, and that low Z points are sometimes higher, and high Z points lower, than the average empirical K/L ratio curve. However, especially for the E2 curve where there seemed to be adequate data, inspection of Goldhaber and Sunyar's curve shows that most examples were of high Z. For this reason their variations are within the bounds of the E2 curves in Figure 1. There was really only one instance, of ^{111}Cd, which was of intermediate Z, and this point was at one end of their curve and therefore did not show a distinct departure. For the M1 curve there is also accumulating evidence that the single curve drawn by Goldhaber and Sunyar is not fully representative. Their line, drawn mainly through points of the tellurium isomer transitions, is probably intermediate between the true curves for elements of low Z like ^{88}Rb, for which there is some evidence of the high M1 K/L ratio of $\sim$8 (Thulin 1952), and curves for high-Z elements like ^{199}Hg, for which the K/L ratio is $\sim$5 (Sherk and Hill 1951).

Comparison of the present curves with those of Hebb and Nelson (1940) and Tralli and Lowen (1949), which were calculated by the approximate method for $Z=35$, further supports the general trend of increasing K/L ratio with decreasing Z.

VI. Conclusion

Both the results presented in this paper and those of Mihelich (1952) indicate that the agreement between the experimental and theoretical L-subshell conversion coefficients may be considered to be very good. Until more complete computations of L-subshell conversion coefficients are available, those already published by Gellman, Griffith, and Stanley (1952) may be used with considerable confidence for confirmation or identification of γ-transition multipolarities.

Further, the L-shell conversion coefficients of Gellman, Griffith, and Stanley may be used in conjunction with the K-shell coefficients of Reitz (1950) or Rose, Goertzel, and Perry (1951) and Rose et al. (1951) to calculate K/L ratios, which may be used in preference to those of Goldhaber and Sunyar (1951), particularly for transitions taking place in the lighter elements.

VII. Acknowledgments

VIII. References

BEACH, L. A., PEACOCK, C. L., and WILKINSON, R. G. (1949).—*Phys. Rev.* **76** : 1585.
BELL, R. E., GRAHAM, R. L., and PETCH, H. E. (1952).—*Canad. J. Phys.* **30** : 35.
CHANG-YUN FAN (1952).—*Phys. Rev.* **87** : 252.
DEUTSCH, M., and WRIGHT, W. E. (1950).—*Phys. Rev.* **77** : 139.
GELLMAN, H., GRIFFITH, B. A., and STANLEY, J. P. (1952).—*Phys. Rev.* **85** : 944.

GOLDHABER, M., and HILL, R. D. (1952).—*Rev. Mod. Phys.* **24** : 179.

GOLDHABER, M., and SUNYAR, A. W. (1951).—*Phys. Rev.* **83** : 906.

GRAHAM, R. L., and BELL, R. E. (1951).—*Phys. Rev.* **84** : 380.

HEBB, M. H., and NELSON, E. (1940).—*Phys. Rev.* **58** : 486.

HEDGRAN, A., and THULIN, S. (1951).—*Phys. Rev.* **81** : 1072.

HILL, R. D., and MIHELICH, J. W. (1950).—*Phys. Rev.* **79** : 275.

KONDAIAH, E. (1951).—*Ark. Fys.* **3** : 47.

METZGER, F. R., and HILL, R. D. (1951).—*Phys. Rev.* **82** : 646.

MIHELICH, J. W. (1952).—*Phys. Rev.* **87** : 646.

MONTALBETTI, R. (1952).—*Canad. J. Phys.* **30** : 660.

MOON, P. B. (1951).—*Proc. Phys. Soc. Lond.* A **64** : 76.

PEACOCK, C. L., and WILKINSON, R. G. (1948).—*Phys. Rev.* **74** : 297.

PLESSET, E. H. (1942).—*Phys. Rev.* **62** : 181.

REITZ, J. R. (1950).—*Phys. Rev.* **77** : 10.

ROSE, M. E., GOERTZEL, G. H., and PERRY, C. L. (1951).—Oak Ridge Nat. Lab. Rep. ORNL-1023 (unpublished).

ROSE. M. E., GOERTZEL, G. H., SPINRAD. B. I., HARR, J., and STRONG, P. (1951).—*Phys. Rev.* **83** : 79.

SAXON, D. (1948).—*Phys. Rev.* **74** : 849.

SHERK, P. M., and HILL, R. D. (1951).—*Phys. Rev.* **83** : 1097.

SIEGBAHN, K., and HEDGRAN, A. (1949).—*Phys. Rev.* **75** : 523.

SLÄTIS, H., and SIEGBAHN, K. (1949a).—*Phys. Rev.* **75** : 318.

SLÄTIS, H., and SIEGBAHN, K. (1949b).—*Ark. Mat. Astr. Fys.* **36A** : 1.

STEFFEN, R. M. (1951).—*Phys. Rev.* **82** : 827.

SWAN, J. B., and HILL, R. D. (1952).—*Phys. Rev.* **88** : 831.

THULIN, S. (1952).—*Ark. Fys.* **4** : 363.

TRALLI, N., and LOWEN, I. S. (1949).—*Phys. Rev.* **76** : 1541.

WEISSKOPF, V. F. (1951).—*Phys. Rev.* **83** : 1073.

WILSON, H. W., and CURRAN, S. C. (1951).—*Phil. Mag.* **42** : 762.

ON THE REACTION ^{23}Na$(p,\gamma)^{24}$Mg AND THE ENERGY LEVELS OF ^{24}Mg

By O. H. Turner*

Summary

The energies of at least 20 γ-rays emitted from the 302 kV resonance level of ^{23}Na$(p,\gamma)^{24}$Mg have been measured with sodium iodide crystals employed as (a) single crystal spectrometer, (b) scintillation pair spectrometer, (c) " cascade " spectrometer. A tentative decay scheme involving levels at $1\cdot38$, $4\cdot25$, $5\cdot3$, $5\cdot8$ or $6\cdot2$, $7\cdot73$, $8\cdot57$, $10\cdot60$, $11\cdot23$, $11\cdot62$, and $12\cdot03$ MeV is proposed for ^{24}Mg. The effect of isotopic spin on the transition probabilities is discussed and some spin and parity assignments made. The resonance concerned has been determined as being at $302\cdot2\pm0\cdot6$ kV, and is believed to be less than 50 V wide.

I. Introduction

Apart from the well-established levels at $1\cdot38$ and $4\cdot14$ MeV that are excited in the decay of ^{24}Na, the energy levels of ^{24}Mg have not received considerable attention. Other levels reported have been (Mandeville 1949) $0\cdot83$, $1\cdot24$, $1\cdot66$, $4\cdot16$, $7\cdot70$, and $8\cdot64$ MeV from the ^{23}Na$(d,n)^{24}$Mg reaction, and $1\cdot5\pm0\cdot2$, $4\cdot2\pm0\cdot2$, $5\cdot5\pm0\cdot3$, $7\cdot3\pm0\cdot3$, $8\cdot3\pm0\cdot4$ MeV from the scattering of 15 MeV protons (Fulbright and Bush 1948). In a more accurate scattering experiment, using 8 MeV protons, Hausman *et al.* (1952) have found levels at $1\cdot38$, $4\cdot13$, and $4\cdot24$ MeV.

The only published measurements of the energies of γ-radiation emitted from the reaction ^{23}Na$(p,\gamma)^{24}$Mg are those of Casson (1953). It would appear worth while supplementing this information with data obtained from experiments undertaken at the Clarendon Laboratory, Oxford, even though the work had to be terminated, some 12 months ago, before conclusive results were reached.

II. Experimental

(a) Accelerator and Target Details

A 1 MV Cockcroft-Walton generator, manufactured by Philips of Eindhoven, delivered a maximum analysed proton current of between about 50 and 150 μA, according to the particular conditions prevailing at the time of operation. The set was ripple suppressed by a phase controlled feedback system that eliminated the fundamental frequency, but left about 100 V peak to peak of second harmonic at 300 kV. As the potential drop across the ion source plasma was about 50 V (Thonemann *et al.* 1948), the beam energy spread was around 140 V at 300 kV. An automatic voltage control kept the mean voltage constant to within 250 V, and careful manual control could maintain the mean voltage within 100 V.

* Physics Department, University of Melbourne.

For resonance measurements, a 3-photoelectric-cell voltage cutter accepted counts within any desired voltage range down to 160 V in width. For a given number of beam integrator counts at a fixed beam potential, counts were recorded from two end-window Geiger counters, arranged so as to eliminate fluctuations in beam position and current distribution within the beam. Thick targets of metallic sodium and calcium fluoride were used.

For the γ-energy measurements, targets of sodium chloride, 50–100 keV thick, were used. About $\frac{1}{8}$ in. of brass and copper intervened between the target and the sodium iodide crystals. To prevent carbon being deposited on the target, a liquid-air cold trap was inserted between the second defining diaphragm and the target chamber, whilst the main silicone oil diffusion pump was fitted with a glycol-refrigerated baffle.

(b) Single Crystal Spectrometer

A conventional scintillation spectrometer was constructed using a 1-in. cube of thalliated sodium iodide, an EMI 5311 photomultiplier of only 35 μA/lm, and either a single channel or a 25-channel pulse-height analyser. The crystal was sealed in a "Perspex" box painted with magnesium oxide and, together with the photomultiplier and cathode follower, was housed in an ebonite tube. Individual points of the single channel runs were made for some multiple of 10,000 monitor counts, the monitoring pulses being taken from the output of the main amplifier. This was considered to be more reliable than the use of a beam integrator, as it automatically corrects for target deterioration or beam scattering. Twelve single channel runs were made, covering the spectrum in three stages, and 18 multi-channel runs were made covering the spectrum above 1 MeV in four stages. In addition two multi-channel runs were made to cover the region below 1·5 MeV in two sections.

(c) Scintillation Pair Spectrometer

The block diagram (Fig. 1) presents the salient features of this spectrometer. The centre crystal was a 1-in. cube, whilst the side crystals were 4 cm in diameter and 3 cm long. A dry box was used to wrap the freshly polished crystals in aluminium foil and then mount them in paraffin-filled thin duralumin containers with "Perspex" windows. The separation of the upper surfaces of the two side crystals was 5·3 cm. The two single channel pulse-height analysers selected pulses in the energy range of 0·3–0·55 MeV. The coincidence unit was usually set to 0·1 μsec resolving time and any coincident pulses opened the cathode follower type of gate with a 100 V, 5 μsec square pulse. Reasonable statistics were obtained for 12 runs varying from 200 to 700 μAh of proton bombardment.

(d) "Cascade" Spectrometer

A single channel pulse-height analyser, selecting a narrow band of the γ-spectrum formed in one crystal, opened a 1·7 μsec gate to allow pulses from a second crystal to operate a 25-channel pulse-height analyser. This enabled energy determinations to be made of the γ-rays in cascade with a selected γ-ray or group of γ-rays. Twenty runs from 100 to 500 μAh were made, the statistics of the shorter runs being rather poor.

(e) Calibration

Calibration was achieved with the use of ^{22}Na, ^{65}Zn, ^{24}Na, ThC″, and ^{9}Be($\alpha,n\gamma$) sources plus ^{19}F(p,γ) and ^{11}B(p,γ) reactions. Linearity was retained for all energies except for an indefinite 0·1–0·2 MeV " shift " at 11·7 MeV when using the pair spectrometer. This shift is attributed to the combined effect of electrons and bremsstrahlung radiation escaping from the crystal. (The apparent position of the pair peak for 17·6 MeV γ-rays was found to have shifted 2·5 MeV for a single 1-in. cube.) The possible presence of ^{19}F contamination was simply checked by comparing runs made above and below the ^{19}F 340 kV resonance.
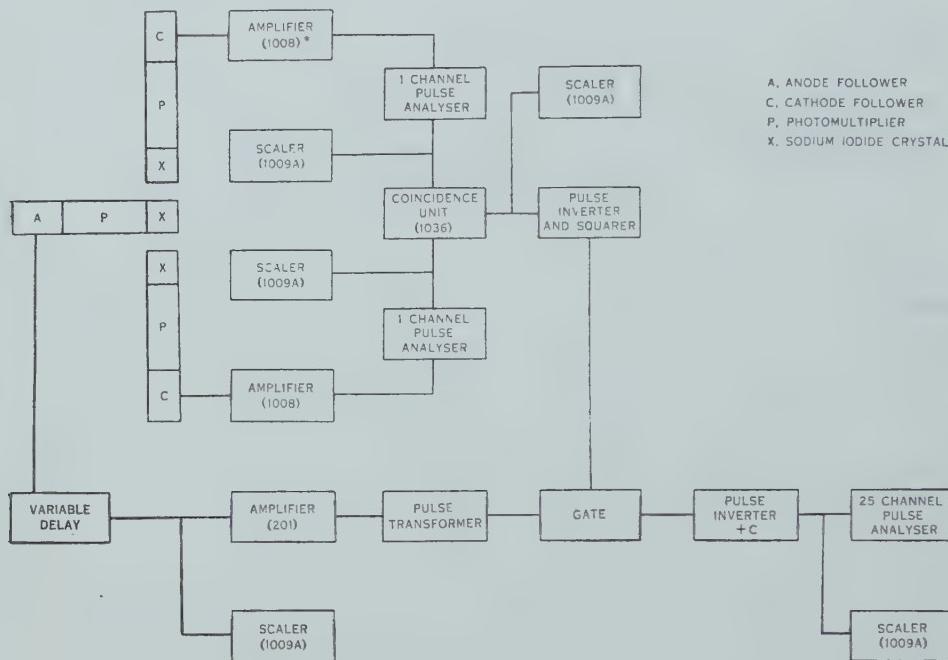
Fig. 1.—Block diagram for scintillation pair spectrometer.
*Electronic unit type numbers correspond to A.E.R.E. (Harwell) specifications.

III. RESULTS

(a) Resonance Values

Assuming that the ^{19}F resonance is at 340·4±0·4 kV (Morrish 1949), then the lowest important ^{23}Na resonance was found to be 302·2±0·6 kV. The apparent width of this level was observed to be even less than the expected experimental width. The full width at half-height of the corrected resonance is believed to be less than 50 V. As the thick-target radiation below 302 kV is less than 1 per cent. of that above that energy, this needle-sharp resonance is ideal for a rapid check on the voltage calibration.

A single rough determination of the next intense level resulted in a value of 594·5±1·5 kV with a full width at half-height (or, more correctly, $\frac{3}{4}-\frac{1}{4}$ of

thick-target radiation increase) of 400 ± 150 V. Burling (1941) determined these resonances as being at 310 ± 10 and 598 ± 10 kV, whilst another level at 676 kV was misprinted as 576 kV. This misprint has been transferred to Alburger and Hafner (1950).

(b) Energy Measurements

The energies, E, of peaks in the spectra obtained from the various experiments are summarized in Table 1. The single crystal data are divided into two groups according as to whether the results were obtained from single channel or 25-

TABLE 1

γ-ENERGY AND INTENSITY

1 Channel		25 Channels		Pairs		Cascade		$E\gamma$ (MeV)	I
E (MeV)	No.	E (MeV)	No.	E (MeV)	No.	E (MeV)	No.		
		0·22	1						~25
		0·41	1					0·41±0·02	~15
0·51	2	0·51	2			0·51	2		~40
						0·63	6	0·63±0·03	≲30
0·81	2	0·8	1U			0·80	6	0·80±0·02	~50
1·0	2	1·0	1			1·1	3	(1·38)	
1·37	2	1·38	3			1·4	12	1·38±0·02	~70
1·85	1U			1·83	3	1·85	6	2·86±0·08	20
2·1	2U	2·1	4			2·2	7U		
2·35	4	2·5	2	2·40	3	2·5	2U	3·43±0·07	14
2·87	3	2·8	3	2·9	6	2·9	1	3·89±0·06	18
3·35	4	3·25	5	3·30	6	3·2	1	4·30±0·06	28
		3·55	7					(4·30)	
4·2	1U	4·25	4	4·3	3U			5·3 ±0·15	1–2
		4·7	3U	4·8	3U			5·8 ±0·15	2–3
5·13	3	5·1	3	5·16	3			6·15±0·10	4
5·73	4	5·65	9	5·75	3			6·75±0·15	3
6·15	3	6·1	6	6·2	3			7·18±0·12	6
6·75	7	6·71	9	6·71	7			7·73±0·06	44
7·4	3U	7·5	2U	7·5	2U			8·5 ±0·2	~1
8·15	5	8·2	4	8·25	4			9·22±0·10	13
8·8	4U	8·85	3U	8·85	3			9·85±0·15	7
9·5	7	9·6	3	9·6	4			10·60±0·12	13
				10·2	2U			11·2 ±0·2	4
10·8	4U			10·8	2			11·8 ±0·15	23

channel pulse-height analysers. The numbers in the column associated with each energy column represent the number of occasions that the particular peak was identified. If the peaks were not statistically significant, the letter U has been added. The energies of the γ-transitions believed to be responsible for the various peaks are listed under E_γ. They have been determined from the mean of all peak energies weighted according to the accuracy of the particular run. The errors are estimates derived from the statistical variation of the apparent peak position, and the extent to which it may be separated from other

peaks.   Casson (1953) reported 11 peaks each of which may be identified with
one of the peaks in Table 1, but his interpretation of some of these peaks differs
slightly from the above.

In each experiment, the spectrum had to be covered in at least three over-
lapping sections, usually with different degrees of amplification.   It is therefore
difficult to combine the various runs into a single curve unless no attempt is
made to insert the original points on the curve.   The pair spectrometer runs are
readily correlated in terms of the number of side crystal coincidences, the total
of which was 132,000.   The amplifications, which varied from 4 analyser
channels/MeV to 11 channels/MeV have been normalized to $5 \cdot 5$ channels/MeV
for presentation in Figure 2.   For comparison of peak shape, four calibration
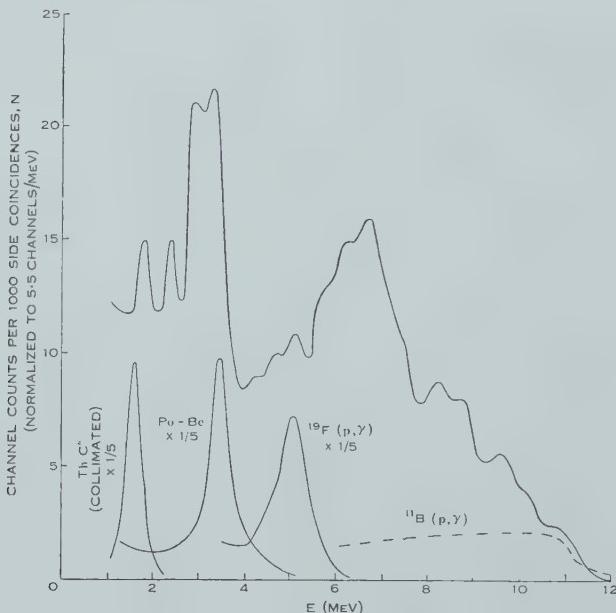
Fig. 2.—γ-Spectrum from ^{23}Na$(p,\gamma)^{24}$Mg, using a pair spectrometer.

energies are also shown.   It is noted that as the γ-ray energy increases, the
change in slope of the lower energy side of the peak is increased, until at 12 MeV
the peak virtually disappears.   This is attributed to the escape of bremsstrahlung
radiation from the crystal.   The rise in the low energy tail of the peaks is
attributed to the escape of electrons from the crystal, plus the detection of
electrons emitted from the target backing.   The pair spectrometer curve for
$11 \cdot 7$ MeV radiation from ^{11}B$(p,\gamma)^{12}$C was found to be inferior to that from a
single crystal, which still exhibited the remnant of a pair peak.   This effect is
considered to be due to bremsstrahlung radiation escaping from the centre
crystal in sufficient quantities to be competitive with annihilation radiation in
operating the coincidence gating circuit, thereby enhancing the detection of
bremsstrahlung events.   As a result, the apparent peak energy appears to be
lowered by about $0 \cdot 2$ MeV for 12 MeV γ-rays.

Using the shape of the calibration curves as a basis for determining the contribution each γ-ray makes to the overall curve of Figure 2, values of $\int_0^E N\mathrm{d}E$ were estimated for each γ-ray, where N represents the number of channel counts per 1000 side coincidences and E the energy in MeV. The intensity was then taken to be $I=(10/\eta)\int_0^E N\mathrm{d}E$ where η is the per cent. efficiency of pair production plus escape of both annihilation quanta for a 1-in. cube of sodium iodide. Intensity values are listed in Table 1. In some instances, such as when two or more γ-rays are insufficiently resolved, these estimates may be in error by a factor of two or more. The intensities of γ-rays with energies less than 1·5 MeV have been estimated from photo-peaks obtained with single crystals. In these cases, η is the per cent. efficiency of photo-electron production plus capture of Compton scattered radiation. The calculated intensity values in this energy range are very approximate.

(c) Energy Levels of ^{24}Mg

As the Q value for the reaction at 302 kV is 12·03 MeV (Alburger and Hafner 1950), energy considerations alone would suggest the simple cascades 10·6+1·4 MeV and 7·7+4·3 MeV. Unfortunately, the results of the cascade experiment do not support such a suggestion. With the gating range set to cover any region between 12·0 and 9·4 (apparent) MeV, the only cascade γ-rays detected were 0·80 and 0·63 MeV. The 0·63 MeV peak is believed to be mainly a photo-peak, and not merely the Compton of the 0·8 MeV γ-ray. The two peaks have the same apparent height.

The 0·83 MeV level of Mandeville (1949) could be responsible for the 0·8 MeV γ-ray if an additional level at either 1·4 (not the 1·38) or 11·4 MeV is postulated. However, all recent proton scattering experiments (Gove and Hedgran 1952 ; Gove and Stoddart 1952 ; Hausman et al. 1952) indicate that the only level below 4·1 MeV is the 1·38 MeV level.

In general, the available data are best represented by assuming that the levels concerned are at 10·60 and 11·23 MeV. In an effort to confirm the existence of an 11·2 MeV level, the proton energy was raised above the 595 kV resonance and a run taken with the gating range set to 9·5–10·5 MeV. It was found that the 0·8 MeV γ-ray was replaced by either a 1·1 MeV γ-ray in the presence of a 1·4 MeV line or otherwise completely to 1·4 MeV. This suggests the possibility that the direct transition to the 1·38 MeV level which appears to be forbidden for the 302 kV resonance is permitted at 595 kV.

With the gating range set to any part of the pair peak energy region of 7·8–9·3 MeV, the cascade spectrum is found to include peaks at 1·4 and 1·1 in addition to 0·8 and 0·6 MeV. This means that the introduction of the 9·2 and 9·8 MeV γ-rays into the gate has added a 1·4 MeV γ-ray to the cascade. Apparently these two γ-rays represent transitions to the 1·38 MeV level from the 10·6 and 11·2 MeV levels respectively.

The next region investigated was in the gating range 6·2–7·5 MeV which introduces the 7·2, 7·7, and 8·5 MeV γ-rays. The cascade peaks were found to

be 1·4, 1·85, 2·9, and 3·2 MeV, with somewhat indefinite subsidiary peaks at 2·2 and 2·5 MeV which could at least partly be accounted for by the contribution from 2·85 and 3·9 MeV γ-rays. The 3·2 MeV peak, which was smaller than the 2·9 MeV peak, may be due entirely to a 3·9 MeV γ-ray, or there may be a contribution from the 4·3 MeV γ-ray. As the cascade spectrum was not investigated below 0·5 MeV, it is also possible that the 0·41 MeV γ-ray was also present. It is tempting to consider the transition 0·4+3·9+7·7 MeV. Such a transition could proceed through levels at about 4·3 and 3·9 MeV, but for reasons similar to those advanced against the existence of 0·8 and 1·7 MeV levels, a 3·9 MeV level is not favoured. However, a level in the vicinity of
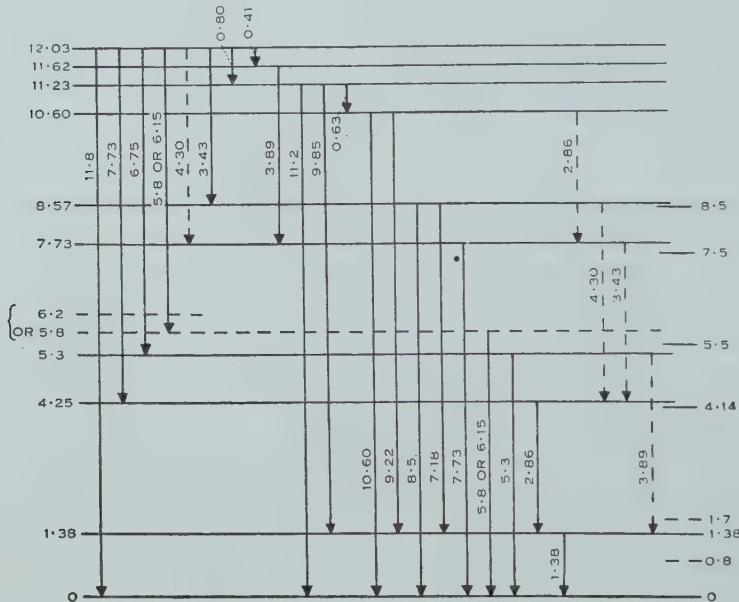
Fig. 3.—Proposed decay scheme for the 302 kV resonance of ^{23}Na$(p,\gamma)^{24}$Mg. Level energies on the right are from Alburger and Hafner (1950). All energies are in MeV.

7·7 MeV has been reported and would suggest that this is the required level, in which case one has to choose between additional levels at either 8·1 or 11·6 MeV. The latter is favoured because (i) it is near the continuum, (ii) it lies in a region not previously investigated, (iii) the 8·1 MeV region has been covered without locating a level.

The other transition involving 7·7 MeV indicated by the cascade results was 7·75+2·85+1·4 MeV. The errors are insufficiently large to suggest that this is in reality a 7·9 MeV γ-ray proceeding to the well-known 4·14 MeV level. Instead, the best fit was for a transition to a level at about 4·25 MeV, which could be a 0+ vibrational level expected from the α-particle model to be at about 4–5 MeV. A level at 4·24±0·02 MeV has since been found by Hausman

et al. (1952). If this level is $0+$, then any $4 \cdot 3$ MeV radiation that may be in cascade with the $7 \cdot 7$ MeV γ-ray would have to be to the $7 \cdot 73$ MeV level.

The $7 \cdot 2$ and $8 \cdot 5$ MeV γ-rays included in the $6 \cdot 2$–$7 \cdot 5$ MeV gating range may be associated with the $2 \cdot 5$ MeV peak which could include the pair peak from the $3 \cdot 43$ MeV γ-ray. As a level is known to exist around $8 \cdot 5$ MeV, it appears reasonable to assume that the transitions $3 \cdot 4 + 8 \cdot 5$ and $3 \cdot 4 + 7 \cdot 2 + 1 \cdot 4$ occur through a level somewhere between $8 \cdot 55$ and $8 \cdot 60$ MeV. As the $5 \cdot 3$ and $6 \cdot 75$ MeV γ-rays add up to the Q value and a level is known to be near $5 \cdot 5$ MeV, a simple cascade through a $5 \cdot 3$ MeV level is assumed. The remaining γ-rays of $5 \cdot 8$ and $6 \cdot 15$ MeV probably also form a direct cascade through a level in the region of $5 \cdot 8$ or $6 \cdot 2$ MeV.

The above conclusions are presented in Figure 3 as the best representation of the existing data. On the right-hand side are the levels given by Alburger and Hafner (1950). The dotted transitions on the right represent possible additional γ-rays whose energies approximately coincide with γ-rays selected in the main decay scheme. In particular, the possible $4 \cdot 3$ MeV transition from the $8 \cdot 57$ to the $4 \cdot 25$ MeV level is wanted to partly fulfil the intensity requirements. Even then the intensity of the $4 \cdot 3$ MeV transition is insufficiently satisfied.

IV. Discussion

Using the formulae for approximate γ-transition probabilities given by Blatt and Weisskopf (1952), even though the formulae are based on a single particle model, a 12 MeV E2 transition is about four times more probable than a $0 \cdot 8$ MeV E1 transition. As the discrepancy between theoretical and experimental lifetimes may be as great as two or three orders of magnitude, it is seen that if the $0 \cdot 4$, $0 \cdot 6$, and $0 \cdot 8$ MeV γ-rays are E1 transitions, then they could successfully compete with the higher energy γ-rays provided that the latter were E2, or higher forbidden, transitions.

It is of interest to invoke the question of isotopic spin. The level of ^{24}Mg corresponding to the ground states of ^{24}Na and ^{24}Al (Birge 1952) is expected to be at about 10 MeV. As the ground state of ^{24}Na has a spin of 4 units (Smith 1951) it is unlikely that the corresponding state in ^{24}Mg is the $10 \cdot 6$ MeV level as it has a direct transition to the ground state. The energy levels of ^{24}Na as determined by Sperduto and Buechner (1952) are compared with the higher levels of ^{24}Mg (Alburger and Hafner 1950) in Figure 4. The position of the ground state of ^{24}Na relative to the ^{24}Mg levels has been deliberately adjusted to indicate the possibility of the $10 \cdot 60$ and $12 \cdot 03$ levels having $T = 1$ and the $11 \cdot 23$ and $11 \cdot 62$ MeV levels having $T = 0$. If this should be true, then the $0 \cdot 4$, $0 \cdot 6$, and $0 \cdot 8$ MeV γ-rays would be unaffected by the isotopic spin selection rule that forbids electric dipole transitions between two pure $T = 0$ states (Radicati 1952). However, in practice there is usually a slight mixing of $T = 0$ and $T = 1$ states (Radicati 1953; Wilkinson and Jones 1953) so that E1 transitions between two nominally $T = 0$ states are inhibited rather than prohibited.

If the $12 \cdot 03$ MeV level is really a $T = 1$ state, then an interesting consequence results. Electric dipole transitions would then be allowed to any level below

10 MeV (being $T=0$ states), but apart from the $1\cdot38$ MeV transition no γ-ray has been found with an intensity considerably greater than the $0\cdot8$ MeV γ-ray, so one is forced to conclude that E1 transitions do not occur for energies greater than $0\cdot8$ MeV. As the ground state, $1\cdot38$, and $4\cdot14$ MeV levels are known to be $0+$, $2+$, and $4+$ respectively, then the $12\cdot03$ MeV level cannot have odd parity (unless $J>5$) and must have even parity. The 302 kV resonance is then
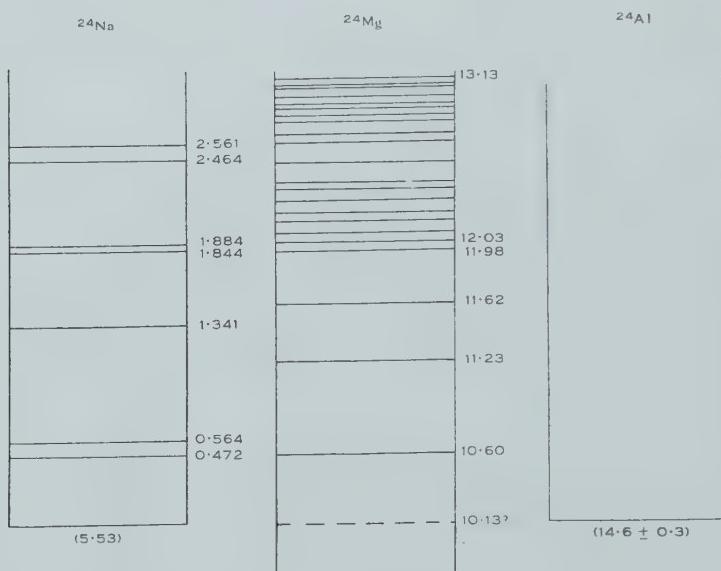
Fig. 4.—Comparison of the known energy levels of the isobaric nuclei
^{24}Na, ^{24}Mg, ^{24}Al over a 3 MeV region.

apparently due to a P proton wave. This agrees with that expected from the narrowness of the level (Christy and Latter 1948). This resonance level may then be a $1+$, $2+$, or $3+$ state. The $3+$ state is unlikely as it would require an M3 transition to the ground state to be much faster than M1, E2 admixtures to the $1\cdot38$ and $4\cdot14$ MeV levels. The possible total angular momentum, J, and parity assignments for all levels are given in Table 2.

TABLE 2

TOTAL ANGULAR MOMENTUM AND PARITY ASSIGNMENTS

E (MeV)	0·0	1·38	4·14	4·25	5·3	5·8	7·73	8·57	10·60	11·23	11·62	12·03
J, parity	0+	2+	4+	0, 1, 2, 3+	1, 2+	1, 2+	1, 2+	1, 2+	1, 2+	1, 2—	0, 1, 2—	1, 2+

Following Wilkinson and Jones (1953), an estimate of the proportion of $T=1$ in the $11\cdot2$ MeV level may be made from the measured intensities of the $11\cdot2$ and $9\cdot8$ MeV γ-rays. If the level were to be $1-$, then the E1 $11\cdot2$ MeV γ-ray has its intensity relative to the E1 $0\cdot63$ MeV γ-ray reduced by a factor of about 40,000 and the E1+M2 $9\cdot8$ MeV γ-ray by about 15,000. If the level

were a 2-state, then the M2 11·2 MeV γ-ray should be one-seventh of the intensity of the E1 0·63 MeV γ-ray, which happens to be about the ratio of experimental intensities.   As about half of the 9·8 MeV γ-rays may be accounted for by an M2 transition, the E1 reduction is of the order of 30,000.   In either case, therefore, about 0·5 per cent. $T=1$ mixed with the nominal $T=0$ state would account for the transition intensities from the 11·2 MeV level.

The validity of the proposed decay scheme appears to be not incompatible with known transition probabilities.   However, it is quite expected that other levels and other transitions are yet to be found, and that some of the transitions suggested are actually incorrect.   There remains, for example, an outstanding difficulty that is not readily explained in terms of spin changes, isotopic or otherwise, and that is the lack of a transition from 12·03 MeV to 1·38 MeV.

Repeating this experiment at higher resonance levels should assist considerably in determining whether the cascades indicated are in the correct sequence. It could also resolve some of the possibly duplicated γ-rays.   The pair spectrometer that was used followed the principle of Johansson (1950, 1952) in having the side crystals larger than the central crystal.   However, it is believed that at high energies a substantial improvement in the resolution of a pair spectrometer would be achieved if the side crystals were smaller than the centre crystal (preferably by about 2 cm).   This would tend to reduce the number of events recorded that involve either electron escape from the crystal or electrons emitted from the target backing.   The effect of bremsstrahlung escape can only be reduced by increasing the separation of the side crystals.   There appears to be an upper limit of usefulness of a pair spectrometer at around 12 MeV.

V. Acknowledgments

VI. References

Alburger, D. E., and Hafner, E. M. (1950).—*Rev. Mod. Phys.* **22** : 373-85.

Birge, A. C. (1952).—*Phys. Rev.* **85** : 753 (A).

Blatt, J., and Weisskopf, V. F. (1952).—" Theoretical Nuclear Physics."   (John Wiley & Sons : New York.)

Burling, E. L. (1941).—*Phys. Rev.* **60** : 340-3.

Casson, H. (1953).—*Phys. Rev.* **89** : 809-30.

Christy, R. F., and Latter, R. (1948).—*Rev. Mod. Phys.* **20** : 185-90.

Fulbright, H. W., and Bush, R. R. (1948).—*Phys. Rev.* **74** : 1323-9.

Gove, H. E., and Hedgran, A. (1952).—*Phys. Rev.* **86** : 574-5.

Gove, H. E., and Stoddart, H. F. (1952).—*Phys. Rev.* **86** : 572-3.

Hausman, H. J., Allen, A. J., Arthur, J. S., Bender, R. S., and McDole, C. J. (1952).— *Phys. Rev.* **88** : 1296-9.

Johansson, S. A. E. (1950).—*Nature* **166** : 794-5.

Johansson S. A. E. (1952).—*Phil. Mag.* **43** : 249-56.

Mandeville, C. E. (1949).—*Phys. Rev.* **76** : 436-7.

MORRISH, A. H. (1949).—*Phys. Rev.* **76** : 1651-7.

RADICATI, L. A. (1952).—*Phys. Rev.* **87** : 521.

RADICATI, L. A. (1953).—*Proc. Phys. Soc. Lond.* A **66** : 139-44.

SMITH, K. F. (1951).—*Nature* **167** : 942-3.

SPERDUTO, A., and BUECHNER, W. W. (1952).—*Phys. Rev.* **88** : 574-9.

THONEMANN, P. C., MOFFATT, J., ROAF, D., and SANDERS, J. H. (1948).—*Proc. Phys. Soc. Lond.* A **61** : 483-5.

WILKINSON, D. H., and JONES, G. A. (1953).—*Phil. Mag.* **44** : 542-7.

ANGULAR DISTRIBUTION OF PHOTOPROTONS FROM NITROGEN

By B. M. Spicer*

Summary

The energy distribution and angular distributions of protons ejected from nitrogen by 11·5 MeV bremsstrahlung have been measured. From the results obtained, it is concluded that the (γ,p) reaction in nitrogen is due to magnetic dipole and electric quadrupole transitions for photon energies of 7·6–11 MeV.

I. INTRODUCTION

The most general feature of photodisintegration experiments is the existence of the so-called " giant resonance " (Baldwin and Klaiber 1948 ; Johns *et al.* 1950). This resonance occurs in the cross section *v.* energy curve of every photonuclear reaction. The mechanism of this reaction was described by Goldhaber and Teller (1948) in terms of electric dipole absorption of γ-rays exciting counter-current motion of proton and neutron " fluids " in the nucleus.

Considerable interest was shown in the (γ,n) reactions in nitrogen, oxygen, and fluorine when the Saskatchewan group showed that, as well as the giant resonance, there was also a smaller resonance on the low energy side of the giant one (Johns *et al.* 1951 ; Horsley, Haslam, and Johns 1952*a*, 1952*b*). It is the nature of the photodisintegration reaction causing this smaller resonance which is of interest here.

Blatt and Weisskopf (1952) support the Goldhaber-Teller postulates regarding the giant resonance, and suggest that the smaller resonance is due to a combination of magnetic dipole and electric quadrupole transitions. They calculate that the relative importance of electric quadrupole to magnetic dipole transitions increases as the square of the quantum energy for the case of a middle-weight element. On this picture, electric dipole absorption becomes important at about 15 MeV, but is negligible at lower energies.

Peaslee (1952) has derived the Breit-Wigner formula for the case where the incident particle is a photon. On the basis of this derivation, he notes that, to produce the large maximum of the giant resonance, electric dipole matrix elements in which the overlap between initial and final states is almost complete are required. Further, he considers that the excited compound state must have a particular structure that corresponds to coherent excitation of the neutrons and protons into counter-current oscillation, in agreement with the Goldhaber-Teller model. The smaller resonance is then assumed to arise from " incoherent " electric dipole excitation—where all neutrons do not move against all protons.

* Physics Department, University of Melbourne.

Electric dipole excitation postulated by Peaslee for the low energy resonance must produce an angular distribution of the form $a + b \sin^2 \theta$, where θ is the angle between the photon beam and direction of the proton. A combination of magnetic dipole and electric quadrupole excitations should give an angular distribution of the form $a + b \sin^2\theta \cos^2\theta$.

The present experiment is an attempt to determine the mechanism of the reaction in the energy region up to $11 \cdot 5$ MeV, by observing the energy distribution and angular distributions of the protons emitted in the $^{14}N(\gamma,p)^{13}C$ reaction. Since the nature of the (γ,p) and (γ,n) reactions is the same, a subsidiary resonance is expected in the cross section v. energy curve of the (γ,p) reaction. It will be more completely separated from the giant resonance in the case of the (γ,p) reaction as the threshold is 3 MeV lower.

The energy distribution and angular distributions of the photoprotons from nitrogen have been measured with accuracy sufficient to discriminate between the two hypotheses mentioned above.

II. EXPERIMENTAL ARRANGEMENT

The source of X-rays was the Melbourne synchrotron, which was operated at $11 \cdot 5$ MeV throughout this work. The energy scale of the machine was calibrated at the thresholds of the photoneutron reactions in ^{109}Ag and ^{63}Cu, which were measured to be $9 \cdot 2 \pm 0 \cdot 2$ and $10 \cdot 8 \pm 0 \cdot 2$ MeV respectively, in agreement with other observations (Baldwin and Koch 1945 ; McElhinney et al. 1949).

The X-rays were collimated into a beam of total angular width of 55 min by a lead collimator, and the X-ray intensity was monitored with an aluminium-walled ionization chamber.

A scattering chamber, similar to that described by Diven and Almy (1950), contained nitrogen at a pressure of 1 atm. The gas served as the target for the X-rays. The protons from the $^{14}N(\gamma,p)$ reaction were detected in Ilford C2 nuclear emulsions (50μ thick), which were placed in the scattering chamber parallel to, and to one side of, the X-ray beam. The 1 atm of nitrogen in the scattering chamber will stop particles from the reactions $^{14}N(\gamma,d)^{12}C$, $^{14}N(\gamma,\alpha)^{10}B$, $^{14}N(\gamma,np)^{12}C$ before they reach the plates. Their thresholds are respectively $10 \cdot 26$, $11 \cdot 7$, and $12 \cdot 5$ MeV.

To make a background run, the X-ray beam was excluded by blocking the collimator opening with a lead plug, all other experimental arrangements remaining unchanged. In this way the effect of neutrons produced in the collimator was estimated. The actual distributions of protons were then found by subtracting the result of the background run from that of the actual run.

A source of background not eliminated by this means is that due to neutrons produced at the synchrotron itself. These neutrons will come almost entirely from the platinum target, for the energy region considered. The number of these neutrons was estimated using the result given by Edwards and MacMillan (1952) for the integrated cross section of the $Pt(\gamma,n)$ reaction. The protons produced in the $^{14}N(n,p)$ reaction induced by these neutrons contribute less than 3 per cent. of the total protons observed.

III. Measurement of Tracks and Treatment of Data

Measurements made on the proton tracks were the projection of the range on to the plane of the plate, the angle to the beam direction, and the final depth in the emulsion reached by the track. To be accepted a track was required to start at the surface of the emulsion and have direction compatible with origin in the irradiated part of the target. The acceptable angular range was 40–140°. Background plates were analysed in the same way as the plates exposed in the actual run.

The energy of the protons at the surface of the emulsion was obtained from the range-energy relation for Ilford C2 emulsions given by El Bedewi (1951). The energy lost by the proton in the gas between target and plate was calculated using the energy loss results given by Aron, Hoffman, and Williams (1949).

$$-\frac{dE}{dx}=\frac{4\pi^2 e^4 z^2}{mv^2}NZ\left[\ln\frac{mv^2}{I}-\ln\,(1-\beta^2)-\beta^2\right]$$

was approximated by

$$-\frac{dE}{dx}=\frac{A}{E}[\ln E+B],$$

where $A=0\cdot0866$, $B=3\cdot253$. This gave a maximum error of less than $\frac{1}{3}$ per cent. for the range of proton energies considered. To obtain the energy of ejection of the proton from the nitrogen nucleus, the following procedure was adopted. If E_1 and E_2 are the energy of ejection from the nitrogen nucleus and the energy of the proton at the surface of the emulsion respectively, the distance travelled in the gas may be obtained from

$$x=\int_{E_1}^{E_2}\frac{dE}{-dE/dx},$$

giving

$$x=\frac{1}{Ae^{2B}}[\text{Ei}\{2\,(\ln E_1+B)\}-\text{Ei}\{2\,(\ln E_2+B)\}],$$

where

$$\text{Ei}(y)=\int_{-\infty}^{y}\frac{e^y dy}{y}.$$

The distance travelled in the gas may also be calculated from the angle of the track to the beam and the known position of the track in the plate. Using these two equations for the distance and knowing the energy of the proton at the surface of the emulsion, the initial proton energy may be calculated.

The uncertainty in energy due to the finite thickness of the target ranged from $\pm0\cdot06$ MeV for protons of 5 MeV emitted at 90° to the X-ray beam, to $\pm0\cdot25$ MeV for $1\cdot5$ MeV protons emitted at 45° to the beam.

To plot angular distributions, the data were grouped into three energy intervals ($2\cdot2$–$2\cdot6$, $2\cdot6$–$3\cdot0$, $3\cdot0$–$4\cdot0$ MeV) and the tracks were grouped into 10° angular intervals, according to the angle they made with the X-ray beam in the laboratory system. The mean differential cross section over the 10°

c

interval was calculated from the number of tracks per 10° interval, by making a correction for the solid angle subtended by each interval at the position of the track. The number of tracks, $N(\theta)$ in a given angular range, found in an area ΔA is given by

$$N(\theta) = \text{const. } d\sigma(\theta) \int_{\text{vol.}} \frac{z}{R} \cdot \frac{\Delta A}{R^2} dV;$$

where z is the height of the centre of the beam above the plate, R is the distance from ΔA to target, and the integration is performed over the volume of gas-target subtending the angle of 10° at ΔA. Carrying through this integration, we find that, in the first approximation,

$$N(\theta) = \text{const. } d\sigma(\theta) \cdot \overline{\sin \theta},$$

where $\overline{\sin \theta}$ is the average value of $\sin \theta$ over the 10° interval.
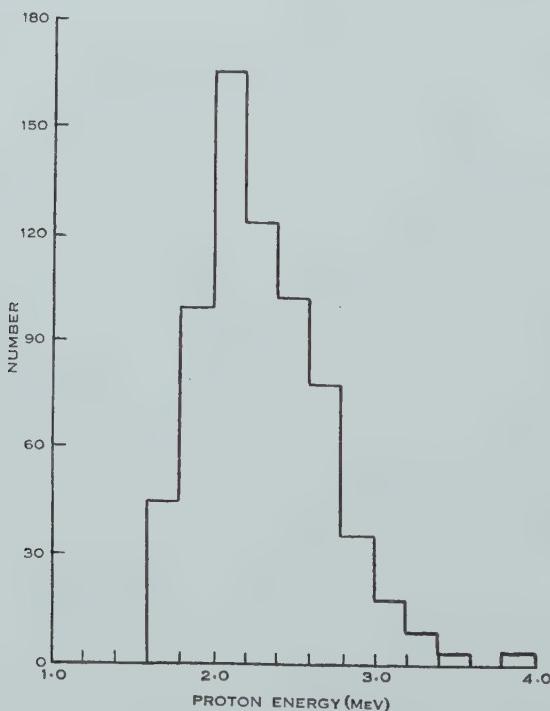
Fig. 1.—Energy distribution of photoprotons from nitrogen.

IV. RESULTS

The results obtained from the measurements are shown in Figures 1 and 2. Measurements were made of 650 tracks.

The energy distribution of protons with energies greater than 2·2 MeV is shown in Figure 1. The lower limit of 2·2 MeV is set because protons of energy less than 2·2 MeV may be stopped in the gas before they reach the plates.

Figure 2 shows the angular distributions plotted for the three energy intervals mentioned above. The angular distribution for the interval $2 \cdot 2$–$2 \cdot 6$ MeV may not be correct as even some of these protons could have been stopped before they reached the emulsion surface, if their angles were less than 60 or greater than 120°. The other two energy intervals provide angular distributions which can be fitted by curves of the form $a + b \sin^2\theta \cos^2\theta$.
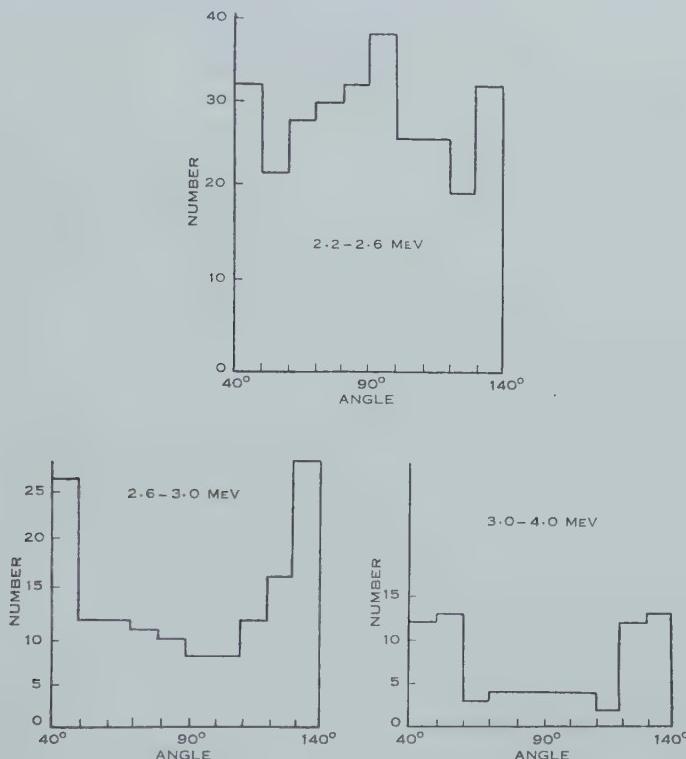
Fig. 2.—Angular distribution of photoprotons.

V. DISCUSSION

The threshold for the reaction $^{14}\mathrm{N}(\gamma,p)^{13}\mathrm{C}$ was calculated from mass tables given by Bethe (1949), Ewald (1951), Li et al. (1951), and Ogata and Matsuda (1953). The mean value found was $7 \cdot 56 \pm 0 \cdot 02$ MeV. Using this value and $11 \cdot 5$ MeV for the bremsstrahlung peak energy, the maximum proton energy expected was $3 \cdot 6$ MeV, calculated from $E_p = 13/14 \cdot (h\nu - E_b)$, where E_b is the threshold for the reaction.

The angular distributions for the proton energy ranges $2 \cdot 6$–$3 \cdot 0$, $3 \cdot 0$–$4 \cdot 0$ MeV can be fitted by curves of the form $a + b \sin^2\theta \cos^2\theta$, where the ratio b/a is greater for the latter case than the former. It is not possible to interpret the angular distribution for protons less than $2 \cdot 6$ MeV in energy. This distribution appears to approach isotropy, when allowance is made for the loss of protons due to their complete stopping in the gas.

The angular distributions for protons in the energy range $2 \cdot 6$–$4 \cdot 0$ MeV are compatible with those expected from a combination of magnetic dipole and electric quadrupole transitions. In this energy range, there is no evidence whatever for the $a + b \sin^2\theta$ distribution expected from the electric dipole absorption postulated by Peaslee (1952).

VI. Conclusion

From the evidence given above the photoprotons from nitrogen are produced by a combination of magnetic dipole and electric quadrupole transitions for quantum energies up to 11 MeV. The magnetic dipole transitions predominate at lower energy, but fall off rapidly in importance. The relative importance of electric quadrupole to magnetic dipole absorption increases with increasing photon energy. Thus, the postulates of Blatt and Weisskopf (1952) are at least qualitatively substantiated.

VII. Acknowledgments

VIII. References

Aron, W. A., Hoffman, B. G., and Williams, F. C. (1949).—Range-energy curves (second revision). U.S. Atomic Energy Commission Rep. A.E.C.U. 663.

Baldwin, G. C., and Klaiber, G. S. (1948).—*Phys. Rev.* **73** : 1156.

Baldwin, G. C., and Koch, H. W. (1945).—*Phys. Rev.* **67** : 1.

Bethe, H. A. (1949).—" Elementary Nuclear Theory." Appendix : Table of Nuclear Species. (John Wiley & Sons : New York.)

Blatt, J. M., and Weisskopf, V. F. (1952).—" Theoretical Nuclear Physics." pp. 651-9. (John Wiley & Sons : New York.)

Diven, B. C., and Almy, G. M. (1950).—*Phys. Rev.* **80** : 407.

Edwards, L. S., and MacMillan, F. A. (1952).—*Phys. Rev.* **87** : 377.

El Bedewi, F. A. (1951).—*Proc. Phys. Soc. Lond.* A **64** : 1079.

Ewald, H. (1951).—*Z. Naturf.* **6A** : 293.

Goldhaber, M., and Teller, E. (1948).—*Phys. Rev.* **74** : 1046.

Horsely, R. J., Haslam, R. N., and Johns, H. E. (1952a).—*Canad. J. Phys.* **30** : 159.

Horsely, R. J., Haslam, R. N., and Johns, H. E. (1952b).—*Phys. Rev.* **87** : 756.

Johns, H. E., Horsely, R. J., Haslam, R. N., and Quinton, A. (1951).—*Phys. Rev.* **84** : 856.

Johns, H. E., Katz, L., Douglas, R. A., and Haslam, R. N. (1950).—*Phys. Rev.* **80** : 1062.

Li, C. W., Whaling, W., Fowler, W. A., and Lauritsen, C. C. (1951).—*Phys. Rev.* **83** : 512.

McElhinney, J., Hanson, A. O., Becker, R. A., Duffield, R. B., and Diven, B. C. (1949).—*Phys. Rev.* **75** : 542.

Ogata, K., and Matsuda, H. (1953).—*Phys. Rev.* **89** : 27.

Peaslee, D. C. (1952).—*Phys. Rev.* **88** : 812.

THE THERMAL AND ELECTRICAL CONDUCTIVITY OF COPPER AT LOW TEMPERATURES

By G. K. White*

Summary

Measurements have been made of the thermal conductivity from 2 to 160 °K and the electrical conductivity from 1·4 to 293 °K of copper in strained and annealed states.

From the results it is concluded that : (i) departures from the additive resistance hypothesis occur for both the thermal and the electrical resistance ; (ii) whereas the values for " ideal " electrical resistivity, assumed due to scattering by lattice vibrations, are well represented by the theoretical relation $\rho_i = CT^5$ $(T < \Theta/10)$, those for the ideal thermal resistivity require an expression $W_i = BT^{2\cdot4}$ $(T < \Theta/5)$ rather than the theoretically derived expression $W_i = BT^2$; (iii) an anomalous increase in the product of the thermal resistance and temperature occurs at low temperatures $(T < 5\ °K)$, which is possibly related to the minimum in the residual electrical resistance.

I. INTRODUCTION

Theory and experiment have shown that, in most pure metals, electrons are almost solely responsible for the conduction of both heat and electricity. Wilson (1936, 1937) and Makinson (1938) have stated that the total resistivity in both cases is the sum of an impurity resistivity (W_0, ρ_0) and an " ideal " resistivity (W_i, ρ_i). Scattering by chemical and physical impurities and scattering by the lattice vibrations are the mechanisms responsible for the respective resistances.

For thermal conduction

$$1/K = W = W_0 + W_i, \quad \dots\dots\dots\dots\dots\dots (1)$$

and for electrical conduction

$$1/\sigma = \rho = \rho_0 + \rho_i. \quad \dots\dots\dots\dots\dots\dots (2)$$

From the experimental and theoretical aspects it is important to know the temperature dependence of these quantities and whether the additive resistance hypothesis is valid. Sondheimer (1950) has shown theoretically that departures from (1) and (2) should give a small increase in the total resistance at temperatures where the impurity and ideal resistances are of the same order of magnitude. As considerable evidence has been collected by previous workers on the failure of Matthiessen's rule for electrical resistance, the work reported here was done in order to compare the departure from additivity for the thermal

* Division of Physics, C.S.I.R.O., University Grounds, Sydney.

resistance with that for the electrical resistance using materials in different states of purity.

Berman and MacDonald (1952) have examined carefully the electrical and thermal conductivities of an annealed specimen of copper of chemical purity similar to that used by the author. Their results confirmed earlier work (cf. Hulm 1950; Andrews, Webber, and Spohr 1951; Berman and MacDonald 1951) on relatively pure metals, which showed that the thermal conductivity at very low temperatures increases proportionately to temperature, reaches a maximum at $T \sim \Theta/15$, then falls approximately as T^{-2} up to $T \sim \Theta/5$, and finally approaches asymptotically a constant value; no trace of the minimum at $\Theta/4$ predicted by Makinson, Wilson, and Sondheimer is observed.

Reference should be made to the comprehensive investigation of Mendelssohn and Rosenberg (1952) on the thermal conductivity of many elements, from groups 1, 2, 3 and the transition group. Their work, however, was generally on one sample only of each element and did not extend above 40 °K, and so did not include a detailed examination of the variation of W or $W-W_0$ with temperature and purity.

In view of the results recently reported on the thermal conductivity of gold (White 1953a) and silver (White 1953b) in different states of purity, it was desirable to extend the measurements to copper and compare them with previous work. Firstly, although the impurity resistivity W_0 appears to vary inversely with temperature, the measurements on gold suggested a small increase in $W_0 T$ at very low temperatures, possibly related to the anomalous increase in ρ_0 that occurs in some metals (cf. de Haas and van den Berg 1936; Mendoza and Thomas 1951); Berman and MacDonald (1952) found a small thermal effect, much smaller than the electrical effect, in their specimen of copper. Secondly, it is well known that the Bloch-Gruneisen formula for the ideal resistivity of a pure metal in which the conduction electrons are quasi-free reduces to

$$\rho_i = CT^5, \quad \text{for } T \ll \Theta. \quad \text{...................... (3)}$$

Similarly, the transport theory predicts that

$$W_i = BT^n, \quad \text{where } n=2 \text{ for } T \ll \Theta. \quad \text{.......... (4)}$$

Measurements on gold by the author gave $n \simeq 2 \cdot 0$, but those on silver indicated that $n \simeq 2 \cdot 4$, hence the desirability of examining copper.

II. Experimental Details

The electrical conductivity of copper has been measured, correct to $\pm 2 \times 10^{-5} \, \Omega$, between 1·4 °K and room temperature in a cryostat (Fig. 1), using a Mueller pattern resistance bridge. The specimen wire S, wound loosely on a mica cross, rests in the inner copper enclosure B, containing a small pressure of helium gas to ensure that temperature equilibrium between the controlled enclosure and the specimen and the gas thermometer D is maintained. This adiabatic shield B is in an evacuated enclosure A, surrounded by liquid oxygen or liquid helium. The temperature of B is maintained by controlling the vapour pressure of liquid oxygen or liquid helium in C, or, in the temperature ranges

4·3–55 °K and above 90 °K, where vapour pressure control is unsuitable, by the electrical heater coil H.   The current to this constantan coil is either varied manually or by an electronic controller (Wylie 1948) activated by the manganin resistance element R. G is a polished copper plate cooled by liquid nitrogen to reduce the heat inflow to the helium bath.   Temperatures are determined from the vapour pressure of liquid in C or from a butyl phthalate filled manometer connected to the helium thermometer D.
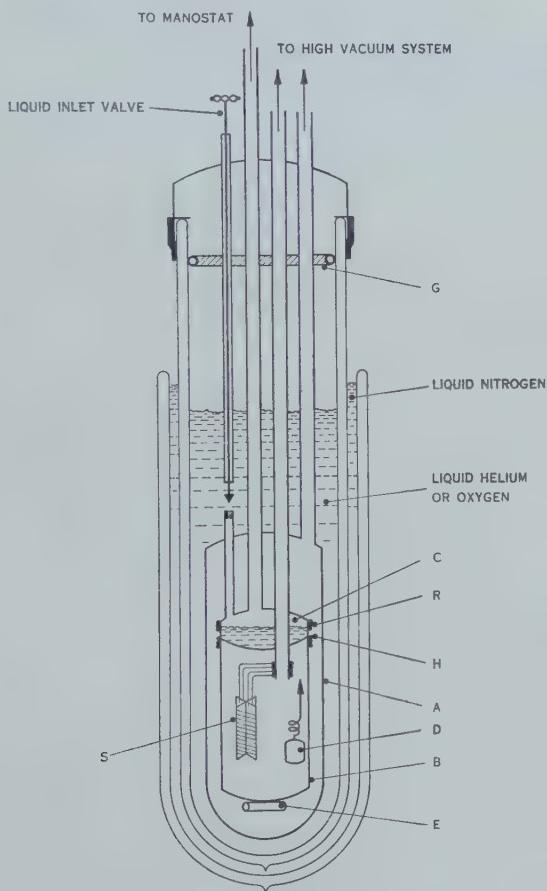
Fig. 1.—Cryostat for electrical resistance measurement.

    The measurements of thermal conductivities have been made between 2 and 160 °K using the apparatus and method described elsewhere (White 1953a). At the higher temperatures a small correction for heat loss by radiation is applied ; the probable experimental error in the results should not then exceed 1 per cent., except between 5 and 15 °K, where high values of conductivity and the necessity for small heat inputs and hence small temperature gradients increases the probable error to a maximum of about 4 per cent.

The measurements reported have been made on specimens prepared from 99·999 per cent. pure copper JM4272 for which Johnson, Matthey and Co. Ltd. quote the following impurities :

$$
\begin{aligned}
&\text{Silver} \quad c.\ 0\cdot0005\%,\\
&\text{Nickel} \quad <0\cdot0003\%,\\
&\text{Lead} \quad\ \ c.\ 0\cdot0004\%,
\end{aligned}
$$

and barely visible spectral lines of gallium and iron.

The specimens for thermal conductivity were as follows :

Cu 1, 2-mm dia. rod drawn by Johnson, Matthey and Co. Ltd., measured in the " as drawn " condition ;

Cu 2, specimen Cu 1 after annealing *in vacuo* at 550 °C for 3 hr ;

Cu 3, 1-mm dia. rod drawn by Johnson, Matthey and Co. Ltd., measured in the " as drawn " condition.
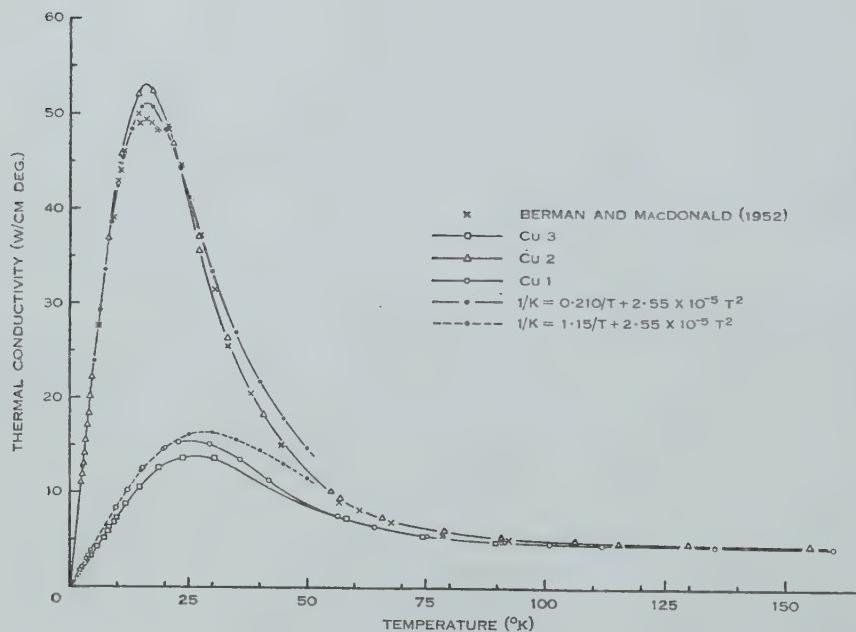
Fig. 2.—Thermal conductivity of copper.

The specimens for electrical conductivity were :

Cu   I, 40 S.W.G. wire drawn by Garrett, Davidson, and Matthey Pty. Ltd., measured in the " as drawn " condition ;

Cu II, specimen Cu I after annealing *in vacuo* at 550 °C for 3 hr ;

Cu III, similar to Cu I.

The rods were about 5 cm long and the wires 300 cm. The wires had a room temperature resistance of about 4 Ω ; since the resistance of the annealed Cu II fell to about 0·05 Ω at liquid helium temperatures, the maximum error in the determination of resistivity is 0·05 per cent.

III. RESULTS

(a) Thermal Conductivity

Figure 2 shows that the thermal conductivity follows the general form discussed above ; as the physical purity is increased the magnitude and temperature of the maximum varies from $13 \cdot 8$ W/cm deg. at 26 °K to 53 W/cm deg. at 16 °K. Disregarding for the moment the small departures of $W_0 T$ from
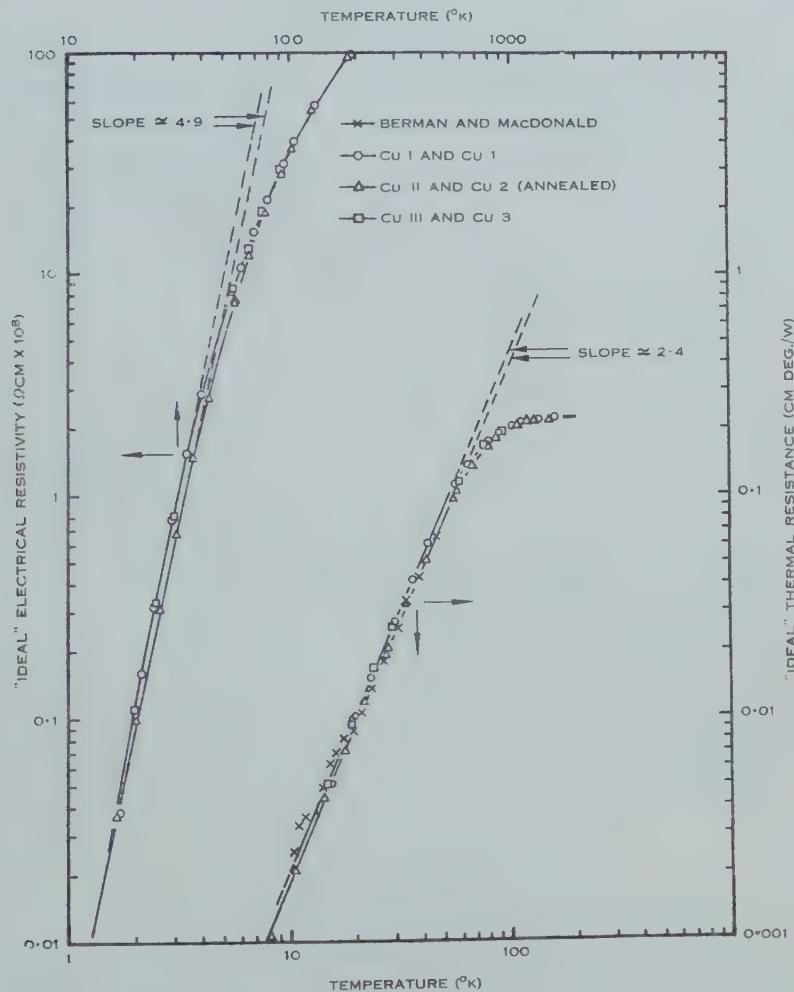
Fig. 3.—" Ideal " electrical and thermal resistance of copper.

constancy, the ideal thermal resistivity W_i will be calculated assuming equation (1) to be valid. The values of $W_0 T$ taken from Figure 4, which are used in calculating W_i for the three specimens Cu 1, Cu 2, and Cu 3, are $1 \cdot 15$, $0 \cdot 210$, and $1 \cdot 32$ cm deg.^{2}/W respectively.

The logarithmic plot (Fig. 3) shows that $W_i = BT^n$ up to a temperature of nearly 55 °K, with $n \simeq 2 \cdot 4$ and with the following values for B :

$$7 \cdot 9_5 \times 10^{-6} \text{ for Cu 1, Cu 3} ;$$
$$7 \cdot 0_6 \times 10^{-6} \text{ for Cu 2}.$$

Figure 4 (a) illustrates the departure of the conductivity from proportionality with temperature in the region where impurity scattering is dominant. The ideal resistivity extrapolated from Figure 3 is so small below 5 °K that $W_i T$ may be neglected in comparison with $W_0 T$, and small corrections have been applied to the values of $W_0 T$ for $T > 5$ °K.
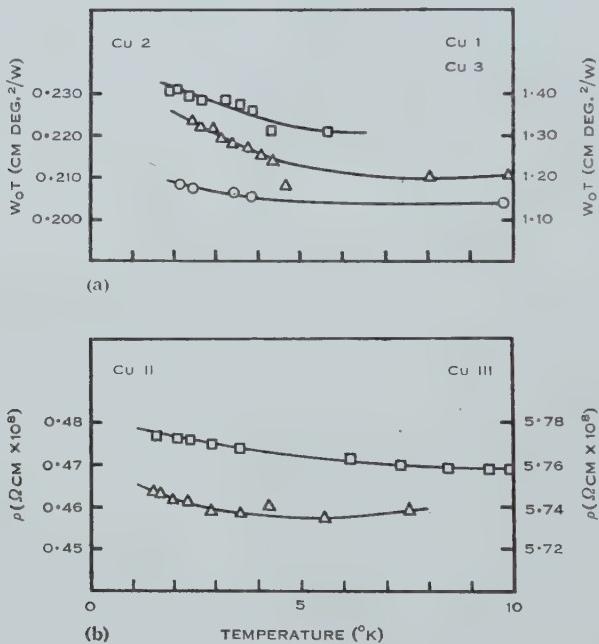
Fig. 4.—Residual resistance of copper below 10 °K. (a) The product of the residual thermal resistance and temperature, (b) the electrical resistance.

—○— Cu 1.
—△— Cu II and Cu 2 (annealed).
—□— Cu III and Cu 3.

(b) Electrical Conductivity

The values of electrical conductivity or resistance are not presented over the full temperature range, as superficially they show no unusual features. The resistance is almost constant below 10 °K and, at relatively high temperatures, increases proportionately to the temperature. For the range from 10 to 35 °K, the resistivity in ohm cm is well represented by

$$\rho = 0 \cdot 0510 \times 10^{-6} + 3 \cdot 7 \times 10^{-16} T^5 \text{ for Cu I,}$$
$$\rho = 0 \cdot 0576 \times 10^{-6} + 3 \cdot 7 \times 10^{-16} T^5 \text{ for Cu III,}$$
$$\rho = 0 \cdot 00458 \times 10^{-6} + 2 \cdot 7_5 \times 10^{-16} T^5 \text{ for Cu II.}$$

The value for C of $2 \cdot 7_5 \times 10^{-16}$ lies reasonably close to the figure of $2 \cdot 64 \times 10^{-16}$ obtained by Berman and MacDonald (1952) for a specimen of similar purity. Figure 3 shows the ideal resistivity ρ_i, calculated from the experimental values of ρ assuming Matthiessen's rule, that is, $\rho_i = \rho - \rho_0$, to be valid.

The small variation in residual resistivity with temperature is illustrated in Figure 4 (b) in which the observed values ρ have been plotted.

IV. Discussion

(a) *Additivity and Ideal Resistance*

In both the electrical and the thermal measurements, the specimens of higher physical purity have a lower apparent ideal resistance, an effect previously noted by the author in the thermal conductivity of gold and silver. This may be accounted for by a departure from the additive hypothesis of equations (**1**) and (**2**). The direction of the departure and the fact that, for the electrical resistance at least, the departure appears greatest in the range 30–40 °K, where the ideal and impurity resistances are comparable, are as predicted by Sondheimer (1950) although the effect appears to be of greater magnitude than the theory predicts. The variation of the ideal resistivity with purity is illustrated by calculation at various temperatures of the Lorenz number $L_i = \rho_i / W_i T$ from the smoothed values of ideal resistivity (Fig. 3) for the annealed and strained specimens. The former yield values in fairly close agreement with those of Berman and MacDonald (1952) ; the latter give values up to 25 per cent. greater at intermediate temperatures ($T \sim 35$ °K) but comparable below 20 °K and above 45 °K.

A feature more difficult to explain is the departure of the temperature dependence of W_i from a T^2 law. It is interesting to note that the results of Berman and MacDonald* on pure annealed copper yield values of W_i (Fig. 3) which appear to vary as T^2 below 20 °K but as $T^{2 \cdot 4}$, in agreement with those of the author, between 20 and 50 °K. Since for copper $\Theta \sim 330$ °K, it is only below $\Theta/10 \sim 33$ °K that the theoretical temperature dependence for W_i might be expected to be well obeyed. In Figure 2 are also shown two curves calculated assuming the relation $W = A/T + BT^2$ to be true, substituting two suitable values of A from these measurements and a value of B reported by Berman and MacDonald (1952).

(b) *Impurity Resistance*

The evidence (see Fig. 4) for an effect in the thermal conductivity related to the changing residual electrical resistance is not conclusive but merits presentation. The electrical resistance of Cu I was not determined with sufficient accuracy to show any obvious minimum, so that an additional series of observations were made on a similar specimen Cu III. The minima observed in both strained and annealed wires were much less marked than that of Berman and MacDonald (1952). However, this is perhaps not surprising, despite the apparent

---

similarity of the samples, in view of the experiments of Gerritsen (1951) and MacDonald (1952) which showed that very small traces of particular elements might produce a strong minimum in the parent metal.

V. Conclusions

The measurements on the thermal conductivity of copper support the theory of non-additivity of the impurity and ideal resistances, as did those on gold and silver reported earlier. The electrical resistance of copper in different states of physical purity indicates a similar departure from Matthiessen's rule in the direction suggested by Sondheimer (1950).

Whereas the ideal electrical resistance varies as T^5 at temperatures below $\Theta/10$, the ideal thermal resistance varies as $T^{2.4}$ up to nearly $\Theta/5$.

At temperatures in the liquid helium region, there is a slight increase in residual electrical resistance, accompanied by a rather similar variation in the thermal parameter W_0T.

VI. Acknowledgments

The author gratefully acknowledges the help of Mr. W. R. G. Kemp and his assistants, who provided liquid helium for these experiments, Dr. P. G. Klemens for much helpful discussion, and Mr. A. K. Sreedhar and Mr. J. V. McAllan for their help in some of the experiments.

VII. References

ANDREWS, F. A., WEBBER, R. T., and SPOHR, D. A. (1951).—*Phys. Rev.* **84** : 994.

BERMAN, R., and MacDONALD, D. K. C. (1951).—*Proc. Roy. Soc.* A **209** : 368.

BERMAN, R., and MacDONALD, D. K. C. (1952).—*Proc. Roy. Soc.* A **211** : 122.

GERRITSEN, A. N. (1951).—Proc. Int. Conf. Low Temp. Phys., Oxford, p. 47.

DE HAAS, W. J., and VAN DEN BERG, G. J. (1936).—*Physica* **3** : 440.

HULM, J. K. (1950).—*Proc. Roy. Soc.* A **204** : 98.

MacDONALD, D. K. C. (1952).—*Phys. Rev.* **88** : 148.

MAKINSON, R. E. B. (1938).—*Proc. Camb. Phil. Soc.* **34** : 474.

MENDELSSOHN, K., and ROSENBERG, H. M. (1952).—*Proc. Phys. Soc. Lond.* A **65** : 385, 388.

MENDOZA, E., and THOMAS, J. G. (1951).—*Phil. Mag.* **42** : 291.

SONDHEIMER, E. H. (1950).—*Proc. Roy. Soc.* A **203** : 75.

WHITE, G. K. (1953a).—*Proc. Phys. Soc. Lond.* A **66** : 559.

WHITE, G. K. (1953b).—*Proc. Phys. Soc. Lond.* A **66** : 844.

WILSON, A. H. (1936).—"Theory of Metals." (Cambridge Univ. Press.)

WILSON, A. H. (1937).—*Proc. Camb. Phil. Soc.* **33** : 371.

WYLIE, R. G. (1948).—C.S.I.R.O. Aust. Div. Phys. Rep. PA-2.

THE SPECIFIC HEAT AND THERMAL CONDUCTIVITY OF GRAPHITE

By P. G. Klemens*

Summary

The specific heat of graphite is discussed in terms of a modified Debye treatment. It is shown that the contribution from the longitudinal waves varies as T^3 below 45 °K, and as T^2 at higher temperatures, whereas the usual two-dimensional treatment leads to a T^2 variation at all low temperatures. Similarly the transverse contribution varies as T^3 at lowest temperatures, but above 10 °K it varies as T^2.

The temperature variation of the thermal conductivity differs from the variation of the specific heat, even though the thermal resistance arises from boundary scattering. This is explained in terms of a mean free path for waves in the hexagonal plane, which is considerably larger for longitudinal than for transverse waves, resulting in an increased contribution of the former to the thermal conductivity.

I. Introduction

The specific heat of graphite at low temperatures is observed to vary as T^n, where according to Gurney (1952) $n \simeq 2$ above 25 °K, and according to Berman (1952), quoting unpublished results of Bergenlid and Hill, $n \simeq 2\cdot2$ from 8 to 20 °K. The thermal conductivity, measured by Berman (1952) for a number of different polycrystalline samples from 3 °K upwards, is observed to vary as T^n, where n ranges from $2\cdot5$ to $2\cdot7$ in the temperature range 3–40 °K. At these low temperatures the thermal resistance is due to scattering of lattice waves by the boundaries of the crystallites.

That the specific heat should vary as T^2 instead of T^3, as is the case for most solids and follows from the usual Debye theory, has been explained by Komatsu and Nagamiya (1951) and by Gurney (1952) in terms of the weak binding between the hexagonal layers. They assume that each layer can be treated separately as a two-dimensional crystal, leading to a T^2 variation. The validity of this treatment will be confirmed, and it will be shown that it breaks down at sufficiently low temperatures.

The thermal conductivity has been interpreted by Berman (1952) in terms of the general theory (Klemens 1951 ; Berman 1953). Heat transport is mainly by lattice waves, and the thermal conductivity of each crystallite, with the resistance arising from boundary scattering, should be of the form

$$\varkappa = \tfrac{1}{3} S v L, \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\text{ (1)}$$

where S is the specific heat per unit volume, v the wave velocity, and L a mean free path determined by the crystal dimensions. The thermal conductivity

* Division of Physics, C.S.I.R.O., University Grounds, Sydney.

should thus have the same temperature dependence as the specific heat. If there are other scattering processes, $\varkappa$ should vary with T more slowly than S, a phenomenon frequently observed with dielectric solids (Berman 1953). Graphite appears to be anomalous, because $\varkappa$ increases with T more rapidly than does S. This can be explained by considering the parts played by the waves of different polarization.

II. Specific Heat of Graphite

Graphite is a layer structure, the atoms in each layer being arranged in a hexagonal array with interatomic distance of $a_0 = 1 \cdot 4$ Å, the separation between these sheets being $a_3 = 3 \cdot 4$ Å. This anisotropy of lattice spacing results from an anisotropy of binding. The bonds between atoms in each hexagonal layer are covalent, reinforced by the bonding of the fourth valency electron, which is partly homopolar and partly metallic. There is thus strong binding in the hexagonal plane, while neighbouring layers are weakly bound by van der Waals forces.

Thus the frequency ω of the lattice waves, a function of the wave-vector $\mathbf{k}$, will depend strongly on k_0, the projection of $\mathbf{k}$ on the hexagonal plane, and only weakly on k_3, the component along the hexagonal axis. In addition, the maximum value of k_3, being $\pi/2a_3$, is smaller than the maximum value of k_0 by a factor of about $2 \cdot 8$.

It will be assumed that ω is given by

$$\omega = c_0 k_0 + c_3 \mid k_3 \mid, \quad \ldots \ldots \ldots \ldots \ldots \ldots (2)$$

where c_0 is considerably larger than c_3. While (2) is not likely to be the correct expression for ω, it does represent the general characteristics of the dependence of ω on $\mathbf{k}$ sufficiently well for the present discussions.

The specific heat per unit volume is given by the following integral over the first zone.

$$S = \sum_j S_j = \sum_j \int \frac{h\omega}{(2\pi)^4} \frac{e^{h\omega/2\pi KT}}{(e^{h\omega/2\pi KT} - 1)^2} \frac{h\omega}{2\pi KT^2} \, d\mathbf{k}, \quad \ldots \ldots (3)$$

the summation being over all polarizations. Consider separately a single mode of polarization. Writing $x = h\omega/2\pi KT$, $y = hc_3 k_3/2\pi KT$, $\theta_0 = hc_0/\sqrt{3} a_0 K$, $\theta_3 = hc_3/4a_3 K$; and noting that

$$d\mathbf{k} = 2\pi k_0 dk_0 dk_3 = 16\pi^4 \frac{K^3 T^3}{h^3 c_0^2 c_3} (x - y) dx dy, \quad \ldots \ldots \ldots (4)$$

S_j can be expressed as

$$S_j = 4\pi \int_0^H dy \int_0^{\theta_0/T} dx K x^2 \frac{e^x}{(e^x - 1)^2} (x - y) \frac{K^3 T^3}{h^3 c_0^2 c_3}, \quad \ldots \ldots (5)$$

where

$$\begin{array}{rll} H = x, & \text{if } x < \theta_3/T, \\ = \theta_3/T, & \text{if } x > \theta_3/T. \end{array} \Bigg\} \quad \ldots \ldots \ldots \ldots (6)$$

In other words, if $f(x)dx$ is the number of normal modes in the (reduced) frequency interval x, dx, then $f(x) \propto x$ for $\theta_3/T < x < \theta_0/T$, as correctly deduced by Komatsu and Nagamiya (1951). But if $x < \theta_3/T$, $f(x) \propto x^2$, whereas these authors, incurring

an algebraic error, stated that $f(x) \propto x$ even at lowest frequencies. We have assumed in (2) that the surfaces of constant frequency are cones or truncated cones, while Komatsu and Nagamiya assume ellipsoids of revolution or truncated ellipsoids. But this should not alter the conclusion derived here that $f(x) \propto x^2$ at lowest frequencies—the only alteration being in the region of transition from x to x^2 dependence.

It follows that for $T > \theta_3$ the important values of x are always larger than the permissible values of y; in that case $x - y \sim x$, and

$$S_j = \frac{4\pi K^4 \theta_3 T^2}{h^3 c_0^2 c_3} \int_0^{\theta_0/T} \frac{x^3 e^x}{(e^x - 1)^2} dx, \quad \dots \dots \dots \dots (7)$$

so that $S_j \propto T^2$ for $\theta_3 < T < \theta_0$. On the other hand, if $T \ll \theta_3$, the limit of the y-integration in (6) is $H = x$, and $S_j \propto T^3$. Since the maximum value of the integrand is in the vicinity of $x = 4$, deviations from the T^3 law will appear above $\theta_3/4$.

The validity of the two-dimensional model, as used by Komatsu and Nagamiya (1951) and by Gurney (1952) is thus confirmed, except for the lowest temperatures, where the T^2 dependence of the low temperature specific heat, derived by these authors, is replaced by a T^3 variation.

The effect of different polarizations must now be considered. For propagation in the hexagonal plane we must separately consider the following three modes: one longitudinal mode, denoted by I, and two transverse modes. Of the latter, one mode, denoted by II, is polarized in the plane, and one, denoted by III, is polarized perpendicular to the plane. It is expected that $c_0^{\mathrm{I}} > c_0^{\mathrm{II}} > c_0^{\mathrm{III}}$. For propagation along the hexagonal axis, modes II and III are equivalent and will be denoted by II; again one expects $c_3^{\mathrm{I}} > c_3^{\mathrm{II}}$, but the ratio of these velocities is different from the corresponding ratio for the 0-direction. For intermediate directions, we assume (2) to hold for each polarization separately. Defining a mean velocity in the 0-direction

$$\frac{3}{<c_0>^2} = \sum_j \left(\frac{1}{c_0^j} \right)^2, \quad \dots \dots \dots \dots \dots (8)$$

the specific heat above the highest value of θ_3 becomes

$$S = \frac{12\pi K^3 T^2}{h^2 2 a_3 <c_0>^2} \int_0^{<\theta_0>/T} \frac{x^3 e^x}{(e^x - 1)^2} dx, \quad \dots \dots \dots (9)$$

and it follows from the observed values of S between 25 and 40 °K that $<\theta_0>$, derived from $<c_0>$, is about 1070 °K. Gurney (1952), attributing the specific heat to mode III only, obtained 614 °K for θ_0^{III}.

Komatsu and Nagamiya (1951) have estimated c_3^{I}, the velocity of longitudinal waves along the hexagonal axis, from the compressibility and, taking account of dispersion, they obtained $\theta_3^{\mathrm{I}} = 180$ °K. Thus S_{I} should vary as T^3 below 45 °K.

In fact the specific heat of graphite seems to follow a T^2 law down to 20 °K, which can be explained only by assuming that c_0^I is appreciably larger than c_0^{III}, so that even around 45 °K S_I is only a small fraction of S. Furthermore, it must be assumed that c_3^{II} is considerably less than c_3^I, so that $\theta_3^{II} \ll \theta_3^I$ and deviations from the T^2 law appear only at much lower temperatures for the transverse component than for the longitudinal component. It appears from the specific heat evidence that θ_3^{II} cannot exceed 35 °K and may well be even appreciably lower, and that above 45 °K $(S - S_I) : S_I = 3 : 1$, so that $c_0^I : c_0^{III}$ is of the order of $3 : 2$.

III. Thermal Conductivity of Graphite

The fact that below 50 °K the thermal conductivity of graphite varies more rapidly with temperature than the specific heat is now explicable in terms of a relatively larger contribution from the longitudinal waves towards the conductivity. We generalize equation (1) to

$$\varkappa = \sum_j \varkappa_j = \sum_j \tfrac{1}{3} S_j c^j L_j, \quad \dots\dots\dots\dots\dots (10)$$

where L is the mean free path due to boundary scattering. In general L is of the order of the shortest linear dimensions of the crystal. This holds even for waves travelling in a direction such that the distance between boundaries is considerably larger, for the lattice waves are coupled together, with respect to their deviation from equilibrium, by three-phonon interactions conserving the total wave-vector. The role of these processes has been discussed elsewhere in detail (Klemens 1951); they tend to equalize mean free paths. Thus the transverse waves in the 0-direction are strongly coupled to waves in the 3-direction. The graphite crystals are in the shape of thin plates parallel to the hexagonal plane. If L_3 is the thickness of these plates and L_0 their diameter, it follows that for transverse waves in the 0-direction $L_{II} = L_3$.

Now it is well known that the lattice waves of the highest velocity cannot interact with any other lattice waves of higher frequency, but only with those of lower frequency. It was shown by the author that such interactions have a negligible effect on the effective mean free path. In graphite the longitudinal waves in the 0-direction have the highest velocity of all, and they are therefore not effectively coupled to any waves in the 3-direction. Consequently for longitudinal waves in the 0-direction $L_I = L_0$.

Since for crystal plates $L_0 > L_3$, it follows that for conduction in the 0-direction the ratio $\varkappa_I : \varkappa_{II}$ is larger than $c_0^I S_I : c_0^{II} S_{II}$, hence larger than $S_I : S_{II}$. Therefore the temperature variation of the thermal conductivity is higher than that of the specific heat.

To account for Berman's data, and provisionally identifying the overall conductivity with the conductivity in the 0-direction, $L_0 : L_3$ must be of the order of 10 : 1, which does not seem unreasonable. It is significant that Berman, by using (1) and taking $v \sim c_0^I$, obtained values for L which were of the order of magnitude of the diameter L_0 of the crystal plates.

For conduction in the 3-direction where $L_I \sim L_{II} \sim L_3$, $\varkappa_I$ is again enhanced relative to S_I, because $c_3^I > c_3^{II}$, the ratio of these velocities from the specific heat

evidence being not less than 5 : 1.   This difference in velocity occurs to a lesser degree in the 0-direction and is unimportant there.   It follows that the temperature variation for conduction in the 3-direction is intermediate between that of specific heat and conduction in the 0-direction.

For crystal plates the directional anisotropy of the conductivity will be of order $L_0 : L_3$.   In a polycrystalline specimen the overall conductivity will be largely governed by the conductivity in the 0-direction.   We have thus a natural explanation for the high temperature dependence of the thermal conductivity of polycrystalline graphite, for the magnitude of the effective mean free path derived from (1), and for the slightly lesser temperature dependence for conduction along the preferred orientation of the C axis.   To explain the temperature dependence of Berman's specimen of smallest crystal size and lowest conductivity, it must be assumed that for these small crystals $L_0 \sim L_3$, so that $\varkappa_I$ is no longer enhanced relative to S_I.

IV. Conclusion

This treatment has been confined to a rough estimate of the relevant parameters occurring in the simplified lattice theory of specific heat and thermal conductivity.   In view of the lack of detailed knowledge of the interatomic forces and the uncertain experimental material, there is no point in refining these calculations.   It does appear, however, that the thermal conductivity can be related to the specific heat according to the usual lattice theory, if account is taken of the special features arising from the high anisotropy of graphite. In order to verify this theory, it would be of value to have measurements of the specific heat from 2 to 8 °K, and of the thermal conductivity of a single crystal of graphite.

V. Acknowledgments

VI. References

Berman, R. (1952).—*Proc. Phys. Soc. Lond.* A **65** : 1029.
Berman, R. (1953).—*Advanc. Phys.* **2** : 103.
Gurney, R. W. (1952).—*Phys. Rev.* **88** : 465.
Klemens, P. G. (1951).—*Proc. Roy. Soc.* A **208** : 108.
Komatsu, K., and Nagamiya, T. (1951).—*J. Phys. Soc. Japan* **6** : **438.**

D

THE EFFECT OF TENSION ON THE THERMOELECTRIC PROPERTIES OF METALS

By A. J. Mortlock*

Summary

Measurements have been made on 11 different metals of the change in thermoelectric power accompanying elastic tensile strain. The measurements were made over the temperature range 20–400 °C on specimens of known purity. There is evidence that the magnitude of the general effect is dependent upon the purity of the specimen. The results obtained with gold differ considerably from those found by another observer.

New approximate values of the general coefficients, which describe the first-order change in thermoelectric power for isotropic metals under all types of elastic strain, have been computed using the results obtained and the only available similar low pressure data. The new values suggest that recent conclusions regarding the position of the Fermi level for gold, silver, and copper require modification.

I. Introduction

It has long been known (e.g. Thomson 1856) that tensile stress produces changes in the thermoelectric properties of metals. This phenomenon has been investigated in some detail by Cohn (1879), Meyer (1896), Maclean (1900), Baedecker and Vehrigs (1914), Bridgman (1918), and Smith (1925). In much of this earlier work the effect of plastic deformation, which may produce effects of the same order of magnitude but sometimes of opposite sign to that produced by elastic deformation (Thomson 1856), was not fully distinguished from the latter (Borelius 1935). More recently Crussard (1948) made quantitative measurements of the change in thermoelectric power caused by pure elastic tensile strain on four pure metals at about 100 °C.

Within the elastic range it is possible to determine the effect of any type of strain on thermoelectric properties, to a first order, from general coefficients. The observed effect is, of course, dependent upon the relative directions of the temperature gradient and strain vectors. The coefficients applicable to isotropic metals may be found by a consideration of the observed effects of longitudinal tension and hydrostatic pressure on suitable metal specimens. Smit (1952) utilizes such coefficients in a discussion of the influence of elastic shear strains on the conductivity and thermoelectric power of cubic metals.

This paper reports measurements of the effect of elastic tensile strain on 11 different metals of known purity over the temperature range 20–400 °C. From the results obtained approximate values of the isotropic general strain coefficients for some of the metals investigated are deduced.

* Division of Physics, C.S.I.R.O., University Grounds, Sydney.

II. Experimental Method

The principle of the method of investigation was the same as Crussard's (1948), measurements being made of the e.m.f. generated by a thermocouple formed from a continuous single metal wire so arranged that one arm was subjected to a longitudinal tension while the other was mechanically free. The apparatus is shown schematically in Figure 1. The cold junction temperature was maintained at 20 ± 1 °C, while the mean accuracy of the hot junction temperature control was approximately ± 4 °C. The e.m.f. was measured with a potentiometer having a least count of $0 \cdot 1$ μV, usually in conjunction with a galvanometer amplifier (Wylie 1951). By this means e.m.f. changes could be measured to about $\pm 0 \cdot 002$ μV.
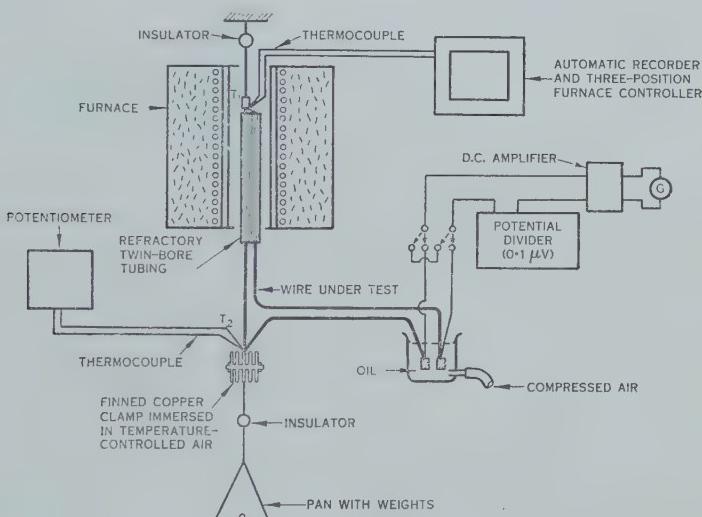
Fig. 1.—Method of measuring a small thermal e.m.f. due to strain.

For each specimen measurements were made at a series of hot junction temperatures, and at each such temperature a definite loading procedure was followed in order to eliminate from the results the effects of plastic deformation which was sometimes present. Measurements were made in succession of the e.m.f.'s E_{W+W_0}, E_{W_0}, E_{2W+W_0}, E_{W+W_0} etc. where W is an incremental load, fixed for any specimen, and W_0 is the weight of the cold junction unit etc. ; the e.m.f. corresponding to a stress intensity $(nW + W_0)$ was taken to be given by

$$\sum_{r=1}^{r=n} (E_{rW+W_0} - E_{(r-1)W+W_0}).$$

This procedure had the advantage over one in which readings were taken as the stress was progressively reduced from the maximum loading used that the measurements were not left incomplete in the event of the wire breaking. Control experiments showed that the e.m.f. obtained by the two methods agreed to within the experimental error, indicating that within the limits of stress intensity used any work hardening of the specimen produced negligible change in the thermal

e.m.f. due to elastic strain. Although any effects due to creep or creep recovery appeared to be negligible, a precautionary period of approximately 5 min was allowed to elapse between readings for reproducibility ; this also allowed the dissipation of heat generated in switch contacts.

The sequence of hot junction temperatures followed for most specimens was 100, 200, 300, 400, 50, 150, 250, 350 °C, thus revealing any change in the characteristics of the specimen with temperature or successive loadings. Measure-

TABLE 1

METALS TESTED

Metal	Origin	Mechanical Condition	Purity or Analysis* (quantities in %)
Copper I	J,M	Annealed	99·999
Silver I	,,	,,	99·999
Silver II	Commercial	,,	~99·94 ; traces Cu, Hg
Gold I	J,M	,,	99·99(+)
Gold II	Commercial	,,	~99·93 ; Ag (0·05), Fe (0·01), trace Cu
Gold III	,,	,,	~99·91 ; Ag (0·06), Fe (0·02), strong trace Zn, trace Cu
Gold IV	,,	,,	~99·90 ; Ag (0·07), Fe (0·02), trace Cu
Platinum	J,M	,,	99·999
Palladium	,,	,,	99·995
Nickel	,,	,,	99·99
Aluminium	Commercial	,,	~99·58 ; Si (0·09), Fe (0·03), Cu (0·01), Mn (0·01), Ti (0·01)
Titanium	,,	,,	Strong traces Fe, Mn, traces Mg, Si
Molybdenum	,,	,,	~99·95 ; trace Si, faint trace to trace Fe
Iron I	,,	(i) Cold drawn ⎫ ⎬ (ii) Annealed ⎭	Undrawn sample ; Mn (0·2), Cu (0·15), Ni (0·1), traces (~0·03) Cr, Sn, As, C
Iron II	,,	(i) As received ⎫ ⎬ (ii) Annealed ⎭	Mn (0·5), Ni (0·1), Cr (0·1), Cu (0·1), C (0·09)
Tungsten	,,	As received	Traces Mo, Si

* Any impurity concentration which was less than or equal to the spectrographic level of " faint trace " has been neglected in the table.

ments were also usually made subsequently at other temperatures, particularly if the earliest readings were inconsistent with those obtained later ; this effect was attributed to straightening of the wire specimens.

III. SPECIMENS

Details of the metals examined, their chemical purity or major impurities or both, and the physical state of the specimens when tested, are given in Table 1. The materials marked J, M were supplied by Johnson, Matthey and Co. Limited, of London, as spectrographically pure. The specimens were annealed by electrical heating *in vacuo*.

IV. Experimental Results

The results are set out graphically in Figures 2 and 3, where the summed e.m.f. increments, E, are plotted as ordinates and the corresponding temperature differences between hot and cold junctions, T_1-T_2, as abscissae ; each curve
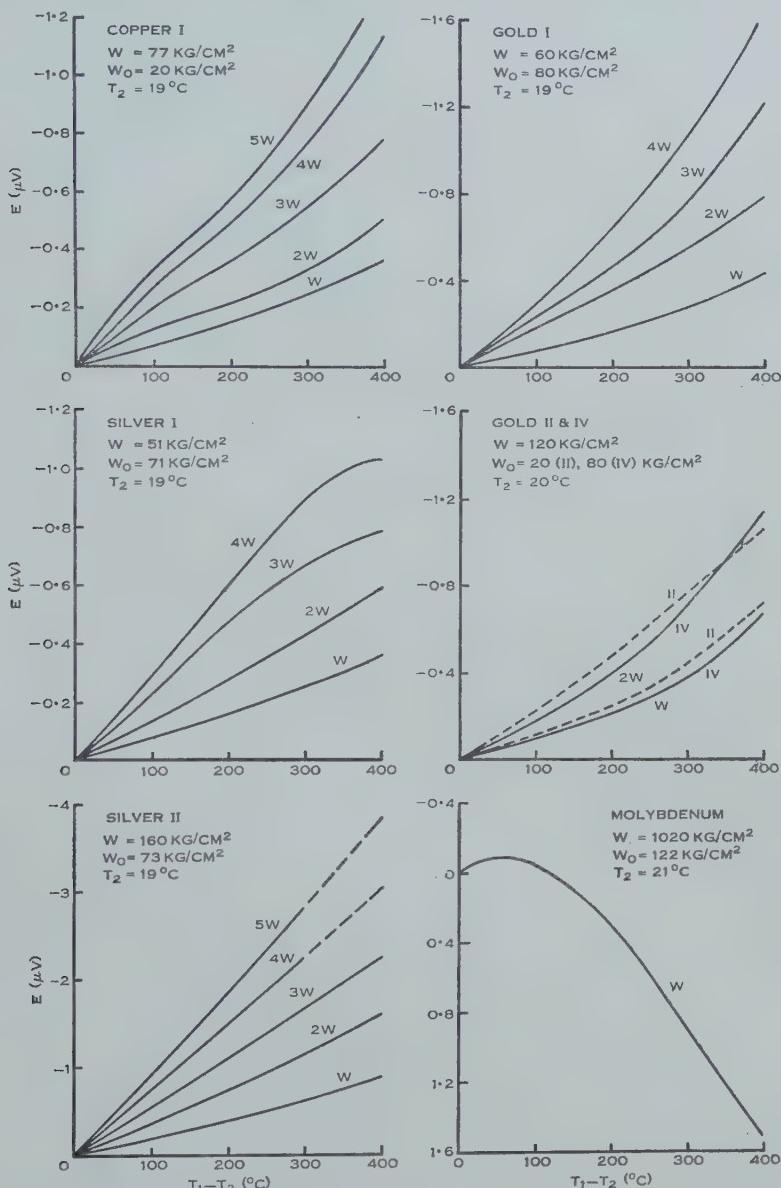
Fig. 2.—E.M.F. of strain thermocouples as a function of the temperature difference between hot and cold junctions for particular stress intensities.

refers to a multiple of the basic stress intensity increment W. The magnitude of both W and W_0 is noted in each figure together with T_2, the mean cold junction temperature.

The sign convention used is that the e.m.f. is taken to be positive if the current flow on completion of the thermoelectric circuit is from the free wire to the stressed wire through the hot junction. The scatter of the plotted points
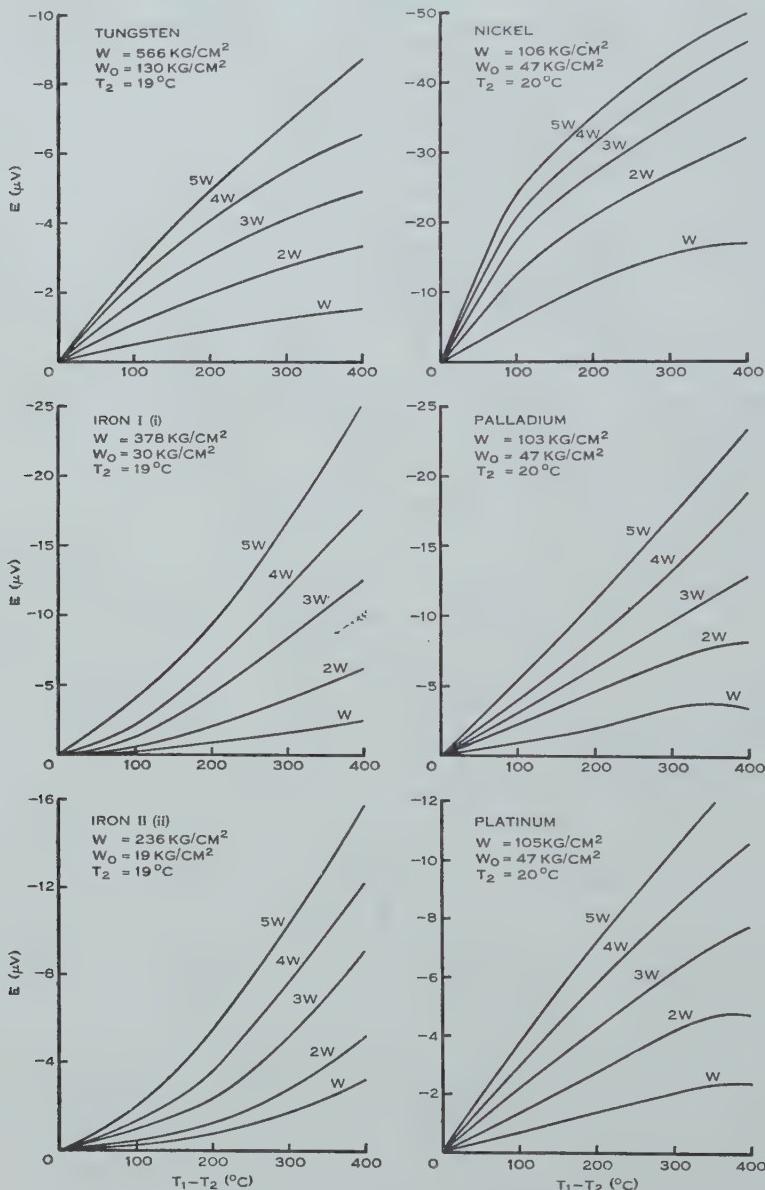
Fig. 3.—E.M.F. of strain thermocouples as a function of the temperature difference between hot and cold junctions for particular stress intensities.

is summarized in Table 2, where the averages of the r.m.s. deviation from each curve are listed. It was found, usually, that the numerical deviation from any given curve was approximately independent of temperature.

Certain of the metals tested require special comment.

Silver II.—The dotted lines in Figure 2 refer to extrapolation into the region where the combination of temperature and stress was sufficient to cause rupture.

Gold.—Plots of the results obtained for gold III are not given as they are similar to those for gold IV. However, the stress range employed was approximately the same in all the gold specimens and the more important results are summarized in Table 3.

Titanium.—The results showed very great scatter, hence it is only possible to give a rough estimate of one of the results for this metal (see Table 3, where $W_0 = 23$ kg/cm² and $t_{max.} = 1440$ kg/cm²).

Molybdenum.—Although a series of readings were taken corresponding to a stress of $W/2$, the ratio of scatter to nominal ordinate was so large that smoothed data can be given for the higher stress only.

Tungsten.—The results given in Figure 3 were obtained on a specimen as supplied, electrical annealing *in vacuo* causing prohibitive embrittlement.

TABLE 2

RANDOM ERRORS OF THE MEASUREMENTS SHOWN IN FIGURES 2 AND 3

Metal	Av. r.m.s. Deviation (μV)	Metal	Av. r.m.s. Deviation, (μV)
Copper I	0·05	Tungsten	0·2
Silver I	0·06	Iron I	0·5
Silver II	0·07	Iron II	0·4
Gold I	0·05	Palladium	1·4
Gold II and IV	0·03	Platinum	0·25
Molybdenum	0·09	Nickel	1·0

Iron.—Before any satisfactory experiments could be carried out on iron wire it was found necessary to make the furnace winding bifilar in order to reduce the internal alternating magnetic field, as any small torsional strain imposed on the test specimen during mounting makes the magnetic susceptibility of the material anisotropic (Dromgoole 1952). In an alternating magnetic field an induced voltage is therefore produced between the ends of the specimen, leading to an unsteady galvanometer spot.

Iron I.—Annealing produced effects which were not reproducible in magnitude, but it is possible to make a few general remarks on the qualitative form of the change.

On annealing, the e.m.f. became predominantly positive and non-linear in its relation to stress, at constant junction temperatures. The e.m.f. generally increased as the load increased, reached a maximum, then decreased. The variation with temperature at constant stress was similar, the e.m.f. increased as the temperature difference increased, reached a relatively poorly defined maximum at roughly 300 centigrade degrees, then decreased.

Iron II.—The curves given in Figure 3 are based on readings obtained on an annealed specimen. Readings taken on an unannealed specimen were practically identical; this was not unexpected in this case since the general mechanical softness of the metal as received suggested that it had already been annealed.

Aluminium.—Measurements on the sample of aluminium were restricted to temperatures between 20 and 150 °C because of the relatively low elastic limit of the material at higher temperatures. It was not found possible to obtain reproducible results, indicating changes in the physical properties of the specimen during the course of the experiments. Such changes are more likely in an impure than a pure specimen.

V. DISCUSSION OF RESULTS

(a) General

It is evident from an inspection of Figures 2 and 3 that the e.m.f. is not linearly related to stress or temperature or both over the full range in the majority of cases. However, an analysis of the scatter of the observations reveals that the e.m.f. for copper, silver, gold, tungsten, palladium, and platinum may in each case be satisfactorily represented as a linear function of temperature (in the range $T_1 = 20$ to 100 °C) and of stress (in the range used).

TABLE 3

TENSION COEFFICIENT OF THERMAL E.M.F. FOR $T_1 = 100$ °C AND $T_2 = 20$ °C

Metal			$\dfrac{(\Delta E)_{\text{av.}}}{80W}$ ($\mu\mu$V/°C per kg/cm^2)	Metal			$\dfrac{(\Delta E)_{\text{av.}}}{80W}$ ($\mu\mu$V/°C per kg/cm^2)
Copper I	..	..	$- 8\cdot9 \pm 0\cdot9$	Titanium	..	..	$-33 \quad \pm 7$
Copper II*	..	..	$- 6\cdot7 \pm 0\cdot5$	Molybdenum		..	$- 0\cdot09 \pm 0\cdot05$
Silver I	..	..	$-13\cdot5 \pm 0\cdot7$	Tungsten	..	..	$-10\cdot0 \pm 0\cdot5$
Silver II	..	..	$-11\cdot4 \pm 0\cdot4$	Iron I	..	..	$-13\cdot5 \pm 0\cdot5$
Gold I	..	..	$-11\cdot8 \pm 1\cdot5$	Iron II	..	..	$-16\cdot3 \pm 1$
Gold II	..	..	$- 9\cdot6 \pm 0\cdot6$	Nickel	..	..	-460 ± 15
Gold III	..	..	$- 7\cdot2 \pm 1\cdot0$	Palladium	..	..	-104 ± 15
Gold IV	..	..	$- 6\cdot8 \pm 0\cdot8$	Platinum	..	..	$- 71 \pm 2$

* The result for copper II has been derived from previous observations (Mortlock 1951) made on commercial wire of unknown purity.

For these metals it is therefore possible to define a single quantity, $(\Delta E)_{\text{av.}}/\{(T_1 - T_2)W\}$, describing the behaviour of the material in this region. Such a quantity may be termed the tension coefficient of thermal e.m.f. and evaluated by averaging the difference, ΔE, between adjacent curves at a value of $T_1 - T_2$ equal to 80 °C.

In Table 3 this tension coefficient of thermal e.m.f., as defined above, is given for each of the metals examined; the errors are estimated standard errors.

Although these data suggest that the general effect is dependent upon the purity of the sample, the evidence is not decisive because of the high relative

errors.   At higher temperatures, however, where the relative errors are smaller, the differences between the tension coefficients evaluated over a common stress range definitely indicate a dependence on purity in each of the four metals so tested.

(b) Comparison with Other Measurements

In Table 4 the measurements previously given by Crussard (1948) and Meyer (1896) for non-ferromagnetic metals are compared with the relevant data obtained in the present investigation.   It should be noted that this comparison assumes the effect to be a linear function over a wider range of stress than that used by the author, and is thus limited in its significance.

TABLE 4

COMPARISON BETWEEN TENSION COEFFICIENTS OF THERMAL E.M.F. FOUND BY VARIOUS OBSERVERS FOR $T_1 \simeq 100$ °C AND $T_2 \simeq 20$ °C

($\mu\mu$V/°C per kg/cm^2)

Metal	Crussard	Meyer	Author
Copper	—7·1 (mean)	— 6·7	— 7·8 (mean)
Silver	—	—10·1	—12·5 (mean)
Gold	—	— 2·3	— 8·9 (mean)
Aluminium ..	+4	+ 8·1	—
Platinum ..	—	—67	—71

Owing to the large discrepancy between the results for gold, the present measurements were extended to a stress intensity equal to that used by Meyer without, however, producing any great change in the tension coefficient.   The discrepancy may be due partly to impurities in Meyer's specimen.   The general agreement otherwise is as good as could be expected.

Ferromagnetic metals are a special case which previous experiments (Meyer 1896) have shown to be sensitive to magnetic state.   The results given in this paper are only included to show a possible temperature variation of the effect, since no special precautions were taken to ensure the reproducibility of the magnetic state.

(c) Relationship between Tension and Pressure Measurements

Assuming small stresses and elastic deformation the change in thermoelectric power $\Delta S_{||}$ parallel to the longitudinal axis of an isotropic polycrystalline wire due to strains $\varepsilon_{||}$ and $\varepsilon_\perp$ respectively parallel and perpendicular to the same axis is given by

$$\Delta S_{||} = \beta_{||}\varepsilon_{||} + 2\beta_\perp \varepsilon_\perp, \quad \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

where the β's are a function of temperature only and the ε's are taken to be positive for contraction.

If the strains are brought about by a longitudinal tension t or a hydrostatic pressure p, this equation may be written respectively as

$$\Delta S_{||} = C_t t \quad \text{or} \quad \Delta S_{||} = C_b p, \quad \ldots\ldots\ldots\ldots\ldots (2)$$

where both t and p are taken as positive and the coefficients C_t and C_b are related to $\beta_{||}$ and $\beta_\perp$ by the following equations:

$$\left.\begin{array}{l} C_t = (2\mu\beta_\perp - \beta_{||})/Y, \\ C_b = (2\beta_\perp + \beta_{||})(1 - 2\mu)/Y, \end{array}\right\} \quad \ldots\ldots\ldots\ldots \quad (3)$$

in which μ and Y are respectively Poisson's ratio and Young's modulus.

Hence if C_t, C_b, and the elastic constants are known for an isotropic metal in a particular state, then the first-order thermoelectric effect of the metal under the same conditions can be calculated for other types of strain.

In experiments of the kind described here $\Delta S_{||}$ is observed directly and is equal to $\partial E/\partial T_1$, T_2 constant. C_t and C_b differ in sign in most cases; this seems to be in agreement with remarks made by Wagner (1908, pp. 996–7), and consistent with the sign of the volume changes, but is at variance with the conclusion reached by Bridgman (1918, p. 376 *et seq.*) and the interpretation of the latter's measurements given in the International Critical Tables (1929).

TABLE 5

THERMAL E.M.F.-STRAIN COEFFICIENTS FOR $T_1 \simeq 100\ °C$ AND $T_2 \simeq 20\ °C$

Metal	$\beta_{||}$ ($\mu V/°C$)	$\beta_\perp$ ($\mu V/°C$)
Gold	11·7	5·6
Silver	14·7 .	6·2
Copper	10·8	1·3
Platinum	132	13·4
Palladium	125 .	6·0

Wagner (1908) has made the only large-scale investigation of the effects of pressure at sufficiently low pressures to be combined with the present tension measurements. Use is therefore made of these data, corresponding to a hot junction temperature of about 100 °C, in order to calculate $\beta_{||}$ and $\beta_\perp$ for those metals which appear to be linear in this region of stress and temperature. In this calculation the unweighted means for C_t of Table 3 and the elastic constants given by Druyvestein (1946) have been used. Values of $\beta_{||}$ and $\beta_\perp$ are shown in Table 5.

(d) Theoretical Consequence of New Result for Gold

It is of interest to consider how Smit's (1952) conclusions concerning the positions of the Fermi surface for gold, silver, and copper may have to be modified, owing to the very large discrepancy between the present result for gold and that found by Meyer.

The author is indebted to P. G. Klemens* for the following comments on this matter: " Smit calculated $\beta_{||} - \beta_\perp$ as a function of the energy gap at the zone boundary (his Fig. 4). To any positive value of $\beta_{||} - \beta_\perp$ corresponded

two values of the energy gap—for one (case (a)) the Fermi surface was inside the zone, for the other (case (b)) the Fermi surface touched the zone boundary. Requiring that the energy gap should be least for copper and largest for gold, and using the experimental values for $\beta_{||} - \beta_{\perp}$ ($0 \cdot 6$ μV/°C for gold, based mainly on Meyer's and Wagner's work), he allocated a position (a) for copper, and (b) for silver and gold, with large differences in the energy gap. But with the new value of $\beta_{||} - \beta_{\perp}$ ($6 \cdot 1$ μV/°C), gold must be moved to a position close to silver, but still case (b). This makes it probable that copper also has an energy gap very nearly the same as for silver and gold, and thus may also fit case (b)."

VI. Acknowledgments

VII. References

BAEDECKER, K., and VEHRIGS, W. (1914).—*Ann. Phys. Lpz.* **44** : 783.

BORELIUS, G. (1935).—" Handbuch der Metallphysik." Ed. G. Masing. Vol. 1, p. 410. (Akademische Verlagsgesellschaft : Leipzig.)

BRIDGMAN, P. W. (1918).—*Proc. Amer. Acad. Arts Sci.* **53** : 374.

COHN, E. (1879).—*Wied. Ann.* **6** : 385.

CRUSSARD, C. (1948).—Report of a conference on strength of solids. Phys. Soc. Lond. p. 119.

DROMGOOLE, W. V. (1952).—*Wireless Eng.* **29** : 115.

DRUYVESTEYN, M. J. (1946).—Philips Res. Rep. **1** : 77.

INTERNATIONAL CRITICAL TABLES (1929).—Vol. 6, p. 225. (McGraw-Hill : New York.)

MACLEAN, F. (1900).—*Proc. Roy. Soc.* **66** : 165.

MEYER, G. S. (1896).—*Wied. Ann.* **59** : 134.

MORTLOCK, A. J. (1951).—M.Sc. Thesis, University of Sydney, p. 34.

SMIT, J. (1952).—*Physica* **18** : 587.

SMITH, H. (1925).—*Proc. Phys. Soc. Lond,* **38** : 1.

THOMSON, W. (1856).—*Phil. Trans.* **146** : 709.

WAGNER, E. (1908).—*Ann. Phys. Lpz.* **27** : 955.

WYLIE, R. G. (1951).—C.S.I.R.O. Aust. Div. Phys. Rep. PA-5.

GALACTIC RADIATION AT RADIO FREQUENCIES

V. THE SEA INTERFEROMETER

By J. G. Bolton* and O. B. Slee*

Summary

The factors involved in the study of discrete sources of galactic noise by the sea interferometer are discussed. Three new forms of sea interferometer which increase the effectiveness of this technique are described.

I. INTRODUCTION

Owing to the low resolving power of metre wavelength aerial systems of reasonable physical dimensions, interference methods have been widely used in the study of radio-frequency emission from the Sun and stars. Two forms of interferometer have been employed, the sea interferometer and the two-aerial interferometer. The former depends on interference between the direct ray from a source and the ray reflected from the sea at an aerial situated on the top of a cliff. This system is analogous to Lloyd's mirror in optics. The other, consisting of two aerials spaced many wavelengths apart along an east-west base line, may be compared to the Michelson interferometer. The relative advantages of the two systems have been previously discussed (Stanley and Slee 1950) but will be briefly reviewed.

One inherent advantage of the sea interferometer is that twice the sensitivity is achieved with a single aerial as with two similar aerials in the other system. Further, no interconnecting cables or preamplifiers are required. Probably the most important advantage is due to the " cut-off " of the sea's horizon. The interference pattern commences sharply as a source rises above the horizon, in contrast to a gradual " fading-in " in the two-aerial interferometer. This feature is most useful in resolving two or more close sources.

Variable refraction adversely affects the sea interferometer in its use for determining accurate source positions and the curvature of the Earth produces effects which restrict its use for measuring angular widths. The two-aerial interferometer is not affected by the Earth's curvature and the effects of atmospheric refraction and scintillation are much smaller than in the sea interferometer. The effects of refraction on sea interference measurements of position can be overcome by taking observations over a sufficiently long period and by calibrating the instrument on sources of known position. Scintillations, which are most severe at low angles of incidence, affect the " seeing " of very weak sources.

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

The best results are obtained near the equinoxes and in the few hours about dawn when scintillations are rarely observed.

The major disadvantage of the sea interferometer is that it is inherently a " total noise " measuring system. The output of the receiver consists of the sum of four components : (1) the noise generated in the receiver itself, (2) the integrated background noise from that part of the sky in the acceptance cone of the aerial, (3) the noise received from the discrete source under observation, and in some cases (4) unwanted signals of terrestrial origin. In certain regions component (2) changes more rapidly than the amplitude of the interference pattern due to the discrete source. The procedure until recently has been to balance out the greater part of the total noise and amplify the remainder in a D.C. amplifier for presentation on a recording milliammeter. It was necessary for the operator to adjust the D.C. amplifier as the change in component (2) from the changing background noise caused the recorder to go off scale. Obtaining records of the discrete sources was thus very laborious and, in addition, the interpretation of the records was often difficult. In the two-aerial interferometer with a phase switching system as described by Ryle (1952) the interference fringe system is switched alternately between two positions half a fringe width apart, a synchronous rectifier being used to detect the difference in receiver output between the two positions. With such an arrangement components (1) and (2) are eliminated from the recorded output and in some cases component (4). Only the difference between the signal from the source in the two positions of the fringe system is recorded and this changes in a sinusoidal manner as the position of the source changes.

Recently several modifications have been made to the simple sea interferometer which reduce or eliminate the effects of the varying background component and greatly increase the sensitivity and ease of operation of the equipment. The first part of this paper gives an account of the physical factors affecting sea interference patterns and the second part describes some of these new techniques.

II. Factors which Govern Sea Interference Patterns
(a) General

When a discrete source, rising above the sea horizon, is observed with an aerial and receiving equipment situated on a high cliff an interference pattern is obtained. The interference occurs as the path difference between the direct ray and the ray reflected from the surface of the sea changes with the altitude of the source. It can be shown that, under idealized conditions, the fringe system for a point source is given by

$$P = 2P_0\left(1 - \cos\frac{4\pi h \sin\alpha}{\lambda}\right),$$

where P is the power received,

  P_0 is the power received with the same aerial in free space,
  h is the height of the aerial above the sea,
  λ is the wavelength,
  α is the altitude of the source.

Interference minima and maxima occur when $4\pi h \sin \alpha/\lambda = 2n\pi$ and $(2n+1)\pi$ respectively, where n is the fringe number. The amplitude of the pattern is given by

$$P_{max} - P_{min} = 4P_0.$$

In actual practice the amplitude is governed by a number of different factors, some of which are due to the geometry of the system and others to the characteristics of the receiving equipment. The expression for the amplitude may be written

$$P_{max} - P_{min} = 4P_0 \, F(z,\alpha) \, . \, F_r(\alpha) \, . \, F_b(n) \, . \, F_c(\alpha,h) \, . \, F_w(n),$$

where $F(z,\alpha)$ represents the aerial sensitivity pattern—a function of altitude and azimuth,

$\quad\quad F_r(\alpha)$ is the factor due to imperfect reflection from the surface of the sea—a function of altitude,

$\quad\quad F_b(n)$ is the factor due to finite bandwidth of the receiver—a function of the fringe number,

$\quad\quad F_c(\alpha,h)$ is the factor due to the divergence in the reflected rays at the curved surface of the Earth,

$\quad\quad F_w(n)$ is a factor due to a combination of the effect of sea waves and the time constant of the recording system.

The various factors may be considered as independent. The factors due to the aerial sensitivity pattern and the receiver bandwidth can obviously be treated separately from the others which are governed by conditions at reflection. The factor due to the reflection coefficient is considered as arising from the physical properties of the sea-water, whereas the curvature factor is due to the divergence of the reflected beam. The curvature factor is only important near grazing incidence, whereas the effect of sea waves is negligible at grazing incidence and important for high fringe numbers.

$F(z,\alpha)$ depends on the aerial used in a particular system and will not be further discussed. Formulae covering the other factors and examples of particular cases will be given in Sections II (b)-(f) following.

(b) The Reflection Coefficient of the Sea

The formulae of Section II (a) assume that the reflection coefficient of the sea is unity. With a reflection coefficient of $r < 1$,

$$P = P_0(1 + r^2 + 2r \cos \Delta),$$

where Δ is the phase difference between the direct and reflected rays.

The maximum and minimum values are $P_0(1+r)^2$ and $P_0(1-r)^2$. Thus the reduction factor due to the reflection coefficient

$$F_r(\alpha) = \frac{P_{max} - P_{min}}{4P_0}$$

$$= r.$$

Theoretical values of the reflection coefficients for vertically and horizontally polarized waves are given by Massachusetts Institute of Technology Radar

School (1946).   For horizontally polarized waves the sea behaves like a perfect reflector and r does not differ sensibly from unity up to altitudes of 10°.   For vertically polarized waves the sea behaves more like a dielectric and the reflection coefficient departs appreciably from unity.   At 100 Mc/s it falls to a value of 0·4 for an angle of incidence corresponding to an altitude of the source of 2·5°.

It is evident that horizontal polarization is to be preferred for sea inter-ferometry, particularly as the phase change on reflection remains constant at π for a wide range of angles of incidence.   Observations on the discrete sources confirm the value of unity for horizontal polarization.   However, the value obtained for the coefficient of reflection for vertically polarized waves is 0·7 instead of 0·4 at its minimum near 2°.
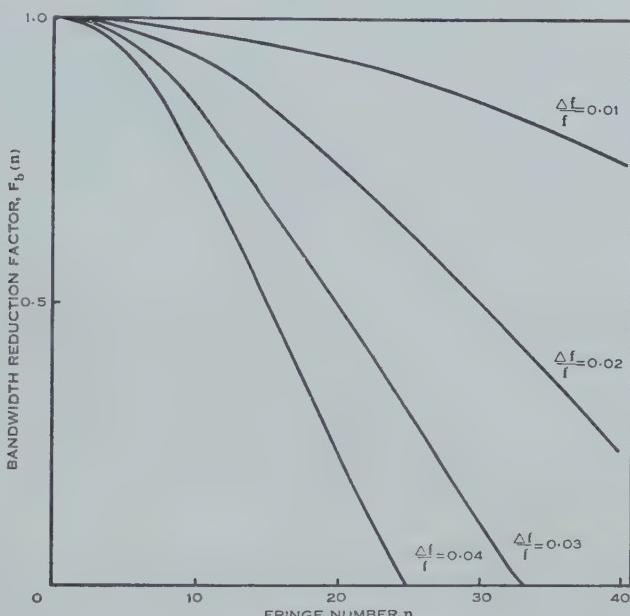
Fig. 1.—Curves showing the effect of receiver bandwidth with increasing fringe number on the amplitude of interference fringes for four ratios of bandwidth to frequency.

(c) The Receiver Bandwidth

The effect of a finite receiver bandwidth on an interference pattern is similar to that in the optical case and results in decreasing fringe visibility with increasing fringe order.   For the sea interference patterns the zero order fringe occurs when the source is first seen on the horizon.   The theory of the effect of a finite receiver bandwidth has been given previously in Part II of this series (Stanley and Slee 1950).   At the nth fringe it can be shown that, for $n < f/\Delta f$,

$$F_b(n) = \frac{\sin n\pi\Delta f/f}{n\pi\Delta f/f},$$

where Δf is the pass band of the receiver (assumed square in shape) and f is the wave frequency.

It will be seen that n and $\Delta f/f$ are interchangeable in the above expression ; thus, for particular values of n and $\Delta f/f$, $F_b(n)$ is the same as that for half the bandwidth and twice the fringe number and so on. Computed curves of the variation of $F_b(n)$ with fringe number for certain ratios of bandwidth to wave frequency are shown in Figure 1. It is clear that when large numbers of fringes are required—for example, for the determination of the position of a discrete source—the receiver bandwidth should be as small as possible. In other applications the use of a wide receiver bandwidth can be advantageous in suppressing unwanted high-order fringes. This has the effect of reducing the acceptance cone of an aerial as far as sources are concerned.

(d) The Curvature of the Earth

Although we speak of direct and reflected " rays ", any aerial has a finite absorption cross section or area and thus receives power from a source in a beam of this cross section. In reflection from a plane surface this beam is not changed but on reflection from a curved surface divergence of the beam occurs. This divergence in sea interferometry results in less power in the reflected beam and hence incomplete interference between it and the direct beam. The theory of the effect of the curved Earth has been developed in Part II of this series (Stanley and Slee 1950) and the amplitudes of the interference pattern at minima and maxima are given by

$$\frac{P_{min}}{2P_0} = \frac{\theta}{\dfrac{h}{R\theta} + \dfrac{3\theta}{2}} \, ,$$

$$\frac{P_{max}}{2P_0} = \frac{\dfrac{2h}{R\theta}}{\dfrac{h}{R\theta} + \dfrac{3\theta}{2}} \, ,$$

where R is the radius of the Earth and θ is the angular separation of the aerial and the point of reflection at the centre of the Earth.

Thus

$$F_c(\alpha) = \frac{P_{max} - P_{min}}{4P_0}$$

$$= \tfrac{1}{2} \cdot \frac{2h - R\theta^2}{R\theta \left(\dfrac{h}{R\theta} + \dfrac{3\theta}{2} \right)} .$$

Since the forms of P_{max} and P_{min} are different, it is more convenient to study the variation of each of these quantities with height of aerial and altitude of source separately. The values of P_{max} and P_{min} are plotted in Figure 2 for four different values of aerial height, which cover the range that could be used in sea interferometry. It can be seen that the effect of the Earth's curvature persists to higher altitudes as the height of the aerial is increased. This limits the usefulness of the sea interferometer in the determination of angular widths of sources. Although the resolving power of the interferometer improves as the

height above the sea is increased and the fringe spacing decreased in consequence, the value of the ratio of the heights of the maxima and minima above the zero level is increasingly affected by the corrections for curvature. It is also clear that for greater aerial heights the sea interferometer loses one of its principal advantages in that the sharp cut-off due to the horizon disappears. The sharp cut-off is replaced by a gradual fading in of the interference pattern (the envelope is indicated by the difference between the curves of P_{max} and P_{min} in Figure 2). This effect is enhanced when atmospheric refraction is taken into account.
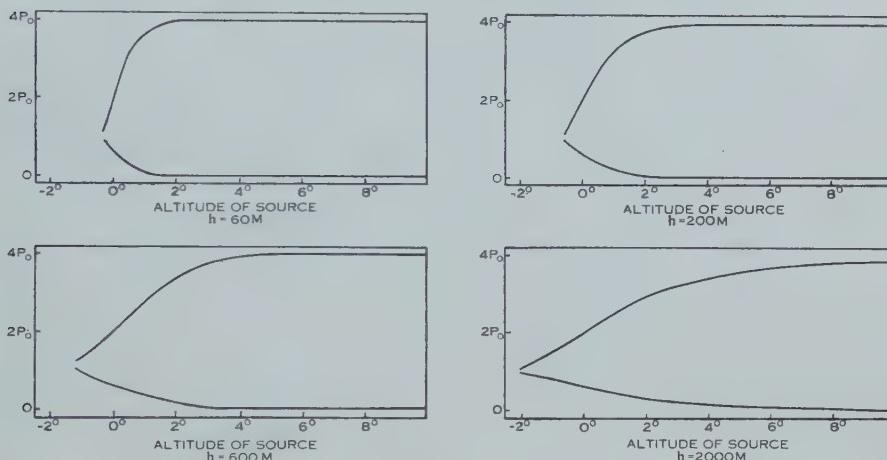
Fig. 2.—Curves showing the effect of the Earth's curvature on sea interference patterns for a range of aerial heights. The upper and lower curves in each diagram represent the variation in the heights of the interference maxima and minima with the altitude of the source. The actual patterns lie within the envelope formed by these curves.

(e) Sea Waves

On a ray theory treatment the area within which reflection takes place for a perfectly smooth sea is the projected area of the aerial system on the sea. This is $H \cot \alpha$ by B, where H and B are the height and breadth of the aerial. In a rough sea not all the radiation incident on this area is reflected towards the aerial and some radiation is reflected towards the aerial from regions outside this area. The phases of the various components are related to the Fresnel pattern on the sea. The dimensions of the first Fresnel zone in and perpendicular to the direction of the source can be shown to be

$$\frac{\lambda \cot \alpha}{\sin \alpha} \quad \text{and} \quad 2\sqrt{\frac{h\lambda}{\sin \alpha}},$$

or

$$\frac{2h \cot \alpha}{n} \quad \text{and} \quad h\sqrt{\frac{8}{n}}$$

in terms of the fringe number. In a typical case, for a cliff height of 80 m and a wavelength of 3 m the dimensions of the first Fresnel zones for fringe numbers

E

2, 10, and 20 are 2300 by 160, 90 by 71, and 23 by 50 m. For an aerial of height 3 m and breadth 10 m the ideal reflection areas for the same fringe numbers are 85 by 10, 17 by 10, and 8 by 10 m.

There are two cases to be considered, one where the dimensions of the waves are small and one where the dimensions of the waves are large compared with the dimensions of the ideal reflection area and the Fresnel zones. The first occurs in a choppy sea and the second in a heavy swell. In the first case phase dispersion occurs in the reflected radiation owing to the dispersion in position of the reflecting regions in relation to the Fresnel zones and owing to the dispersion in heights. Both dispersions increase with the altitude of the source and the phase dispersion due to position may be quite considerable as the aerial beam width includes a number of zones. The effect of the phase dispersion is incomplete interference between the direct and reflected radiation and a decreasing visibility of the fringes with altitude. Observing experience has shown that a choppy sea has little effect on the visibility of the first 10 or so fringes so that for these fringes reflection must mainly occur within the first Fresnel zone.

On the other hand a heavy swell (of wavelength of the order of 100 m) has a marked effect on relatively low-order fringes. It appears in this case that a considerable part of the reflection area rises and falls as the successive crests and troughs cross the line of sight. The effect of this is to produce a secondary modulation of the basic interference pattern with the frequency of the sea waves, as illustrated in Figure 3. This effect on the low-order fringes is most marked
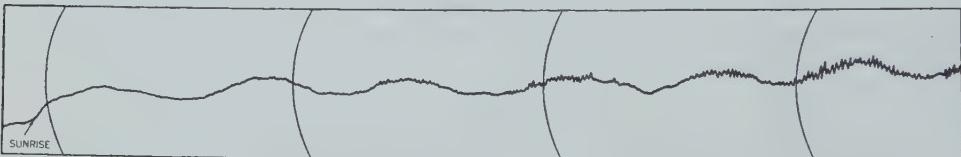
Fig. 3.—Sea interference pattern of the Sun rising showing the increasing effect of sea waves with fringe number. The curved scale lines represent intervals of 10 min. The slow sinusoidal variation is the interference pattern and the ripples are due to sea waves.

in the case of a swell crossing the line of sight of the source, probably because the reflection takes place mainly within the first Fresnel zone which includes part of only one wave. The modulation of the basic interference pattern by the waves is of an unusual type; the deflexions near the interference minima are positive and those near the maxima negative. Thus, if a large time constant is used at the output of the receiver to damp out the modulation, the amplitude of the pattern is reduced compared with that of an unmodulated pattern. The reduction in fringe visibility, which increases with the fringe number, is similar to that due to a finite receiver bandwidth. As the high-order fringes can be reduced by decreasing the beam width of the aerial in the vertical plane, both these effects may be considered as improving the resolving power of the aerial in observations on sources.

(f) Observing Experience and Typical Records

The present writers have used sea interference technique to study the variation of the flux density with frequency of a few major sources with relatively small aerials and in the search for large numbers of sources with a large single-

40 Mc/s JULY 22, 1949

60 Mc/s JUNE 13, 1949

85 Mc/s JUNE 13, 1949

100 Mc/s JUNE 13, 1949

190 Mc/s AUGUST 2, 1950

270 Mc/s SEPTEMBER 5, 1950

400 Mc/s JUNE 22, 1950
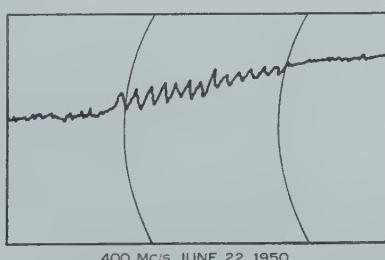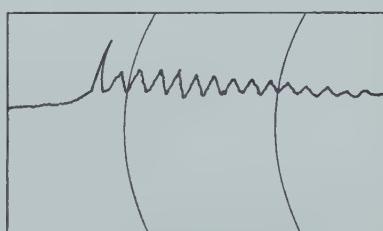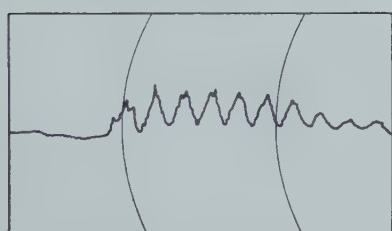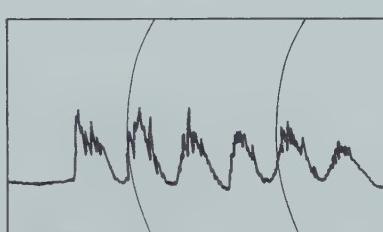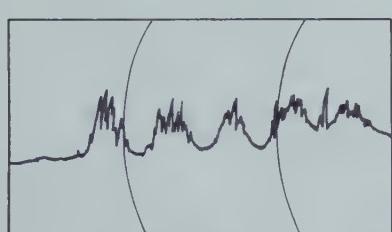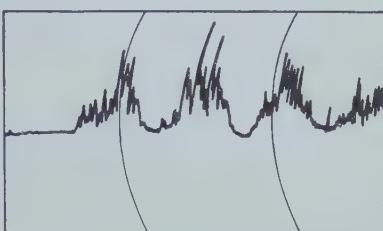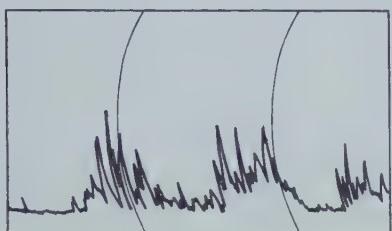
Fig. 4.—Sea interference records of the Cygnus source at rising at various frequencies between 40 and 400 Mc/s.   Note the severity of the scintillations at 40 Mc/s and also the similarity of the scintillations on the 60, 85, and 100 Mc/s records which were obtained on the same day.

frequency aerial.   The records shown in Figure 4 are examples of sea interference patterns of the Cygnus source at rising in the frequency range of 40–400 Mc/s.   The records show that successful observations can be made in

this 10 : 1 range of frequencies.   They also illustrate the variation of the scintilla-
tions with frequency, for, although not all the records were obtained on the same
day, they are representative of the average behaviour on the different frequencies.
At 100 Mc/s and above, reasonable estimates of the flux density of a source can
be made at any time of the day or year.   At 60 Mc/s accurate measurements
are restricted to periods of low scintillation, while at 40 Mc/s useful observations
are rarely obtained due to extreme scintillations and variations in atmospheric
refraction.

Most of our measurements have been made from a height of 80 m although
some observations have been made at heights ranging from 30 to 300 m.   At
heights of more than 300 m the effect of the Earth's curvature becomes
increasingly important and this feature must always be considered in the
selection of a site for observations.

No measurements have been made above 400 Mc/s.   At our present site,
and at this frequency the fringe period for a source of zero declination is only
1 min.   This puts an upper limit of about 10 sec on the size of the time constant
that can be used in the output of the receiver.   The size of this time constant
is one of the factors which determine the ultimate sensitivity of the receiving
equipment, and the permissible time constant becomes increasingly important
at high frequencies due to the rapid decrease in the flux densities of the sources
with increasing frequency.   However, with a lower site we consider that the
sea interference technique could be used successfully for the observation of
sources at frequencies as high as 1000 Mc/s.

III. New Forms of the Sea Interferometer
(a) Systems for Reducing the Effect of the Background Noise

In general there are two possible ways of reducing or eliminating the effect
of the varying background noise in an interference technique.   The first is to
use some form of electronic control in one or more stages of the receiver.   The
second depends on switching either the aerial beam or the interference fringes
between different positions and detecting the difference in signals.   Ryle (1952)
has exploited the fringe switching (or phase switching) method in his two-aerial
interferometer, but this system is not directly applicable to the sea interfero-
meter as the positions of the fringes are determined solely by the geometry of the
instrument.   Three systems which have been used for reducing the effect of the
background noise will be described in Sections III (b)-(d) following.

(b) The Beam Switching System

In this technique the aerial beam is directed alternately 25 times a second
towards the horizon and towards an adjacent region of the sky away from the
horizon.   The aerial consists of two banks (Fig. 5) about $1\frac{1}{2}$ wavelengths apart.
Connection to the receiver is made alternately at the mid point of the cable
joining the two banks and a point $\frac{1}{4}$ wavelength away.   With the banks con-
nected in phase, the aerial beam is directed towards the horizon and the aerial
receives the background noise in this direction and the signal from a source
within the beam.   With the banks connected out of phase, the aerial beam is
directed away from the horizon and receives the background noise from an

adjacent region of the sky.   In the out-of-phase connection the beam is actually split, the lower half being reflected from the sea, but a combination of various factors such as the roughness of the sea and a wide receiver bandwidth ensures that there is no interference within the beam.

The alternating component in the receiver output due to the beam switching is detected by a rectifier synchronized with the aerial switch.   (This type of detector is subsequently referred to as a " synchronous rectifier ", which is suggested as a more suitable alternative to " phase sensitive rectifier " used by other writers.)   The efficiency of the beam switching system depends on the difference in levels between the background noise in the two positions of the switched beam.   It has been found to be very effective in regions away from the galactic plane and also near the galactic plane between longitudes 300 and 60°, where the plane of the Galaxy rises almost perpendicular to the horizon.
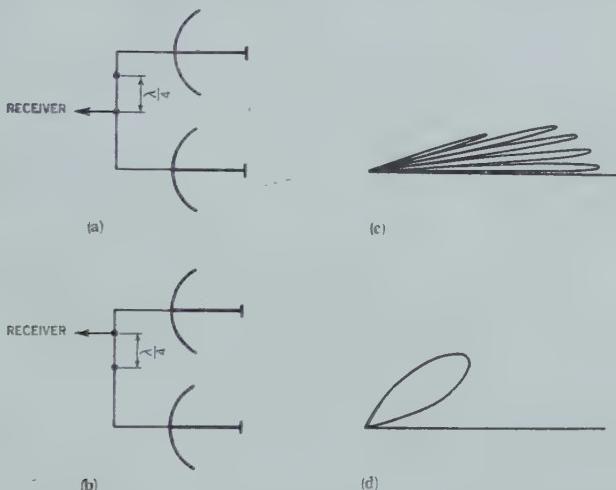
Fig. 5.—Illustrating the principle of the beam switching system for reducing the effect of the variation in background noise. (a) Aerials connected in phase ; (b) aerials connected out of phase ; (c) aerial beam directed towards horizon ; (d) aerial beam directed away from horizon.

The beam switching system greatly reduces the effects of natural and man-made electrical disturbances as long as the source of the disturbance is not located entirely within one of the positions of the switched beams.   It also offers the very useful possibility of being able to detect weak sources close to much stronger ones.   With the sea interferometer weak sources can be detected until such time as a much stronger one rises above the horizon.   A weak source rising shortly after a stronger one can sometimes be detected by the beating of the two interference patterns, but not when the ratio of the flux densities exceeds about 10 : 1.   In the beam switching system, the in-phase and out-of-phase aerial diagrams overlap to some extent and thus at some altitude there is a null point in the switched pattern.   This null point can be changed by changing

the lengths of the connecting cables. By using hand-operated or motor-driven variable lines it would be possible to keep a strong source in the null point after it had attained an altitude of, say, 3° and detect a weaker source at its rising.

(c) Automatic Control of the Receiver Gain

In this system the variations in the background noise are electronically suppressed. The principle of the system is shown in Figure 6. The upper diagram represents the normal output of the receiver, which is fed into an integrator. The output of the integrator roughly follows the input except that rapidly changing components such as the sinusoidal voltages due to discrete sources are smoothed out. The output of the integrator is then used to control the gain of some stage of the receiver, giving a resultant receiver output shown in the lower diagram of Figure 6.
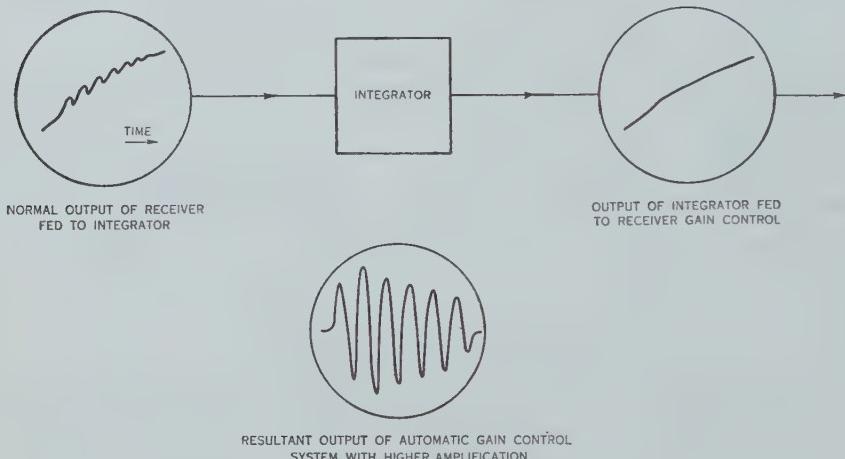
Fig. 6—Illustrating the principle of the automatic gain control system.

The integrator in use utilizes the " Miller " effect. It has three stages of highly stable D.C. amplification giving an overall gain of about 1500. In combination with a 1 MΩ input resistance and a feedback condenser of 2 μF this gives a time constant of about 40 min. The requirements of the integrator time constant depend on the degree of control of the background noise necessary for the section of the sky under observation. Greater control can be obtained by the use of short time constants but a time constant which is too short attenuates the interference patterns. It has been found that with a time constant of 40 min the variation in the background noise is reduced by a factor of 50–100, while the amplitudes of the interference patterns (5–10 min fringe periods, depending on the declinations of the source) are reduced by less than 5 per cent.*

The effectiveness of this system is illustrated in Figure 7 which shows records of the same region of the sky using the total noise system and the automatic gain control system.

---

* The system produces a small phase shift on the fringes. Correction for this is necessary in accurate measurements of the position of a source.

A block diagram of the units of the automatic gain control system in present use is shown in Figure 8. A push-pull output from the integrator is used to control the gains of the two radio-frequency stages in an electronic switch, one of which is connected to the aerial and the other to a reference load. The electronic switch and synchronous rectifier are components of a rapid inter-comparison system introduced by Dicke (1946), but are not essential requirements
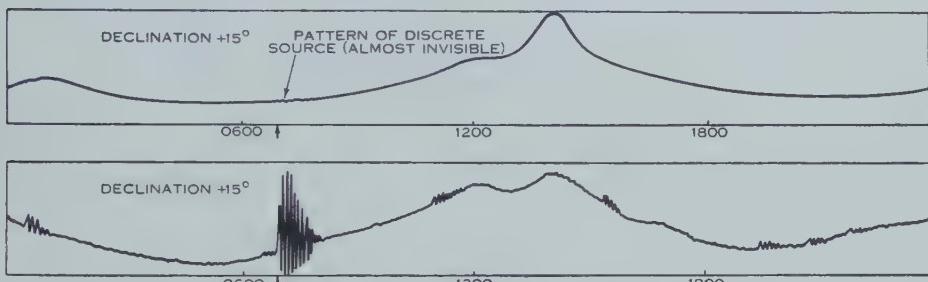
Fig. 7.—Records demonstrating the effectiveness of the automatic gain control system. The upper record is a normal total noise record and the lower is one of the same region obtained using the automatic gain control system. Note that the amplitudes of the interference patterns due to discrete sources are increased by a factor of about a hundred while the swing due to variations in the background noise is about the same.

of the equipment, as the control voltage from the integrator could be applied at a number of other places in the receiver. The inclusion of the switch prevents unwanted receiver gain variations affecting the operation of the integrator and permits a combination of the beam switching and automatic gain control systems to be used when required.
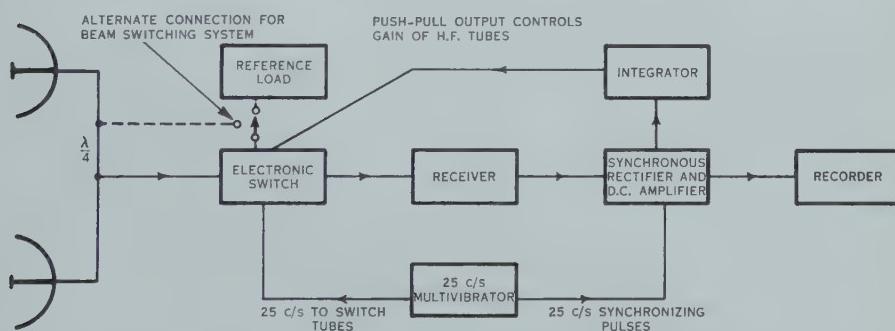
Fig. 8.—Block diagram of the automatic gain control system.

(d) The Azimuth Interferometer

In this system two aerials spaced along a cliff edge are used. The spaced aerials produce a second set of interference fringes at right angles to the normal sea interference fringes. The two aerials are connected to the receiver alternately in and out of phase, thus switching the azimuth fringe system through half a fringe width. A synchronous rectifier at the receiver output detects only

signals due to sources or irregularities in the background distribution smaller than the angular separation of the azimuth fringes. Complete elimination of the background noise can be achieved if sufficiently large aerial spacings are used.

The stages in the development of the system now in use are illustrated in the idealized records of a source rising shown in Figure 9. Figure 9 (a) is a standard sea interference pattern superimposed on a varying background level, as is obtained with a single aerial. Figure 9 (b) results from the combination of two aerials separated by a distance less than twice the cliff height. Phase switching in the two-aerial system eliminates the background noise giving the pattern of Figure 9 (c). However, the time taken for the generation of the azimuth fringes leads to difficulties in the interpretation of the results, particularly when there is more than one source in the aerial beam. This difficulty is
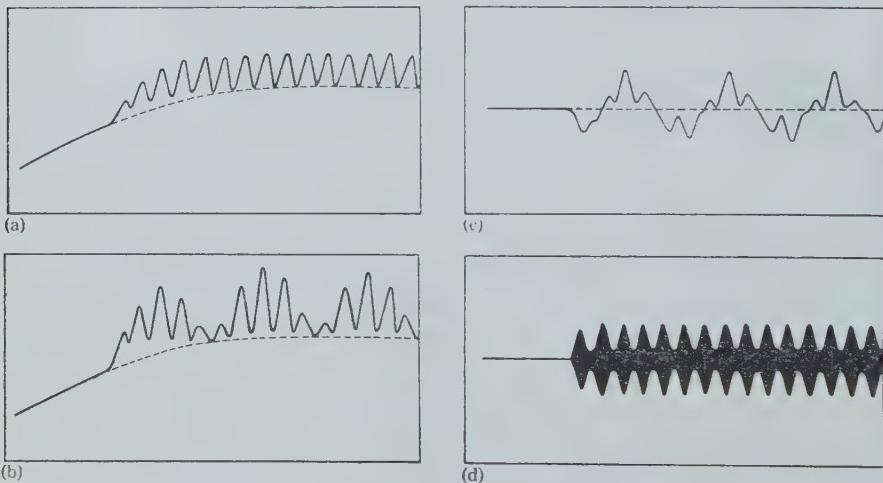
Fig. 9.—Diagrams illustrating the development of the azimuth interferometer. (a) Single aerial, total noise system ; (b) two aerials, total noise system ; (c) two aerials, phase switching system ; (d) two aerials, phase switching and beam swinging.

overcome in a further development (Fig. 9 (d)) in which the azimuth fringe system is swung rapidly backwards and forwards through one fringe width. Here the envelope of the pattern is filled in by the rapid movement of the recorder pen.

A schematic diagram of the equipment is shown in Figure 10. The fringe swinging is achieved by varying the relative phase of the local oscillator to the two mixers by means of variable length lines. Artificial lines are used consisting of fixed inductances and variable capacitances which are driven by a small electric motor. The saturated amplifiers are to prevent the changing impedance of the lines dragging the frequency of the local oscillator or affecting the mixer stages. The phase switching is carried out at the intermediate frequency by means of quarter- and half-wavelength cables and an electronic switch.

Elimination of the background noise in the azimuth interferometer means that use can be made of relatively high-order fringes in making measurements of the angular sizes of sources.   To obtain high resolving power observations have to be made from high cliffs where imperfect reflection from the Earth's curved surface affects the first few fringes.   In the total noise system, measurements on the high-order fringes require an accurate knowledge of the background level.   Its removal by the azimuth system overcomes this difficulty.
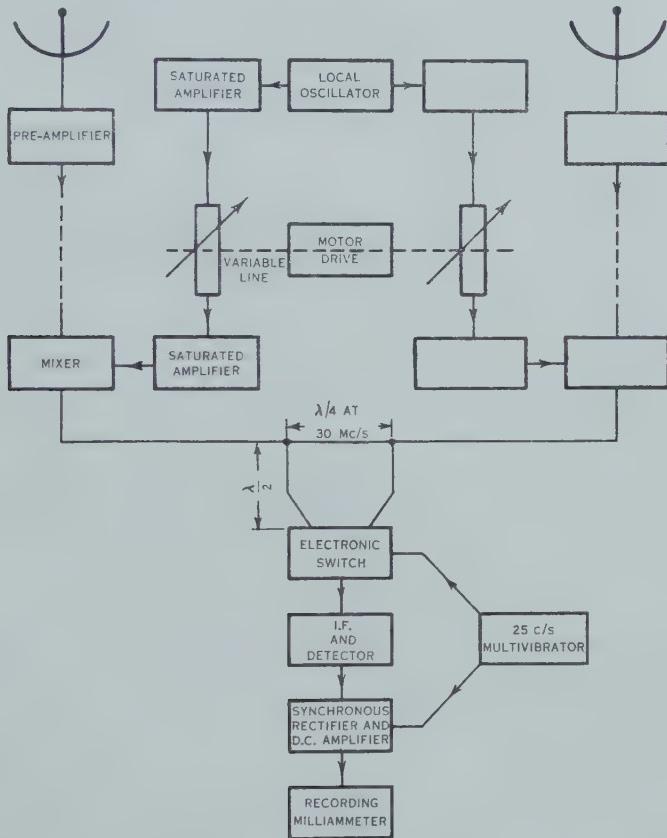
Fig. 10.—Block diagram of the units of the azimuth interferometer.

The azimuth interferometer with its dual fringe system has been found to be of considerable value in the study of extended sources of galactic noise (Bolton 1952).   These observations and observations with the other forms of the sea interferometer will be described in subsequent papers.

IV. REFERENCES

BOLTON, J. G. (1952).—Extended sources of galactic noise.   U.R.S.I. Report 1952.

DICKE, R. H. (1946).—*Rev. Sci. Instrum.* **17** : 268-75.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY RADAR SCHOOL (1946).—" Principles of Radar."
        2nd Ed.   (McGraw-Hill : New York.)

RYLE, M. (1952).—*Proc. Roy. Soc.* A **211** : 351-75.

STANLEY, G. J., and SLEE, O. B. (1950).—*Aust. J. Sci. Res.* A **3** : 234-50.

GALACTIC RADIATION AT RADIO FREQUENCIES

VI. LOW ALTITUDE SCINTILLATIONS OF THE DISCRETE SOURCES

By J. G. Bolton,* O. B. Slee,* and G. J. Stanley*

Summary

A study has been made of the scintillations of four discrete sources at altitudes of from 0 to 10°. The observations cover the years 1947–1951 and were made at various frequencies in the range 40–300 Mc/s. It was found that the scintillation index, a measure of the amplitude of the scintillations, (1) increases with increasing wavelength, (2) decreases rapidly with increasing altitude, (3) shows seasonal and diurnal variations, the seasonal component having minima near the equinoxes and the diurnal component near dawn and sunset. The scintillation rate or the number of scintillations per minute (1) is different for sources of different declination, (2) is independent of wavelength. In the case of the Cygnus source, the data for which are the most extensive, the rate increases and the decline in the scintillation index with altitude is less rapid during the winter months.

A strong correlation is established between the occurrence of the scintillations and sporadic E. The difference in the scintillation rates for different sources can be explained in terms of variations in the size of irregularities and the effects of the winds in the E_s layer.

I. INTRODUCTION

In 1946, Hey, Parsons, and Phillips (1946) observed short period fluctuations in the intensity of galactic noise from a small region in the constellation of Cygnus. Later, Bolton and Stanley (1947) showed that these fluctuations came from what was effectively a point source and they discovered a number of other point sources. At first it was believed that the fluctuations were inherent in the sources, but since then spaced receiver observations have shown that the fluctuations are caused in some region of the Earth's atmosphere.

Observations with spaced receivers, one at Sydney and the other in New Zealand, were first made by the present writers in 1948. Sea interference technique was used at both sites for the observations of the source in Cygnus and the periods of observation overlapped for an hour each night. No correlation was found between the individual scintillations—as they are now generally called—and on some occasions no scintillations were observed at one site and violent scintillations at the other. Control observations were made of solar radiation during a period of high solar activity and almost 100 per cent. correlation was found between solar bursts and other short period fluctuations in the intensity of the solar radiation at the two sites.

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

From these observations it appeared certain that the greater part of the scintillation phenomena was due to atmospheric effects. There remained, however, the possibility of a small fluctuating component in the radiation from the source itself, and this was not discounted until the experiments of Lovell and Little (1950) and Smith (1950). These were carried out over much smaller base lines and with the source almost overhead. They showed that the correlation between individual scintillations at two sites disappeared for distances greater than about 5 km. A later but less extensive series of observations in Australia, some in conjunction with Mr. B. Y. Mills of this Laboratory (Mills and Thomas 1951), gave much the same result.

During the past 5 years about 2000 observations have been made of the four bright sources in Cygnus, Virgo, Taurus, and Centaurus. Each observation covers a period of about 2 hr from the rising of the source above the sea horizon. Most of the observations were made on a frequency of 100 Mc/s but a number were made at various frequencies in the range of 40–400 Mc/s. All observations subsequently described refer to a frequency of 100 Mc/s, unless stated otherwise. The first part of this paper gives a detailed account of the results of an analysis of this material and the second half establishes a correlation between the scintillations and sporadic E.

II. Scintillation Observations
(a) Definition of Terms

Before proceeding with a description of the scintillation phenomena it is necessary to define certain terms which will be used throughout the rest of this paper. It is believed that the flux densities of most of the discrete sources remain constant over long periods. On certain occasions, however, the amplitude of the signal received from the source fluctuates about a mean level with a period of about 1 min. Typical records of these fluctuations are shown in Figure 1. Here the slow variations in the signal are due to the interference method of observation. It can be seen that the amplitude and duration of the individual scintillations are by no means constant, so that any indices describing them are necessarily rough.

We shall define a *scintillation index* as the ratio of the mean peak to trough variations of the signal amplitude to its mean amplitude and express it as a percentage. For example, the scintillation index for the record of Figure 1 (*a*) approaches 200 per cent. at 1635 hr, and the mean index for the whole record is about 70 per cent. ; the mean index for the record of Figure 1 (*b*) is between 30 and 40 per cent. Estimates by eye were used in assessing the records.

We shall define a *scintillation rate* as the number of peaks per minute. In determining scintillation rates from the records, where small scintillations of the order of fluctuations in the receiver noise occurred, every alternate small scintillation was counted.

(b) Seasonal and Diurnal Variations in the Scintillation Index

From 1947 to 1951 observations were made of the source in Cygnus at its rising almost every second day. Each record was assessed to give a daily scintillation index and the daily indices averaged to provide monthly means.

The values for the years 1948, 1949, and 1950 are shown in Figure 2 (*a*).   The curves in this figure show a pronounced apparent seasonal variation in the scintillation index.   It may be only apparent, as a diurnal variation could produce such a curve due to the change in the local time of observation of the source of 24 hr during the year.

All the available data for the four sources in Cygnus, Virgo, Taurus, and Centaurus are shown in Table 1.   Some of the monthly means for the fainter sources are based on so few observations that the reliability of the mean index is
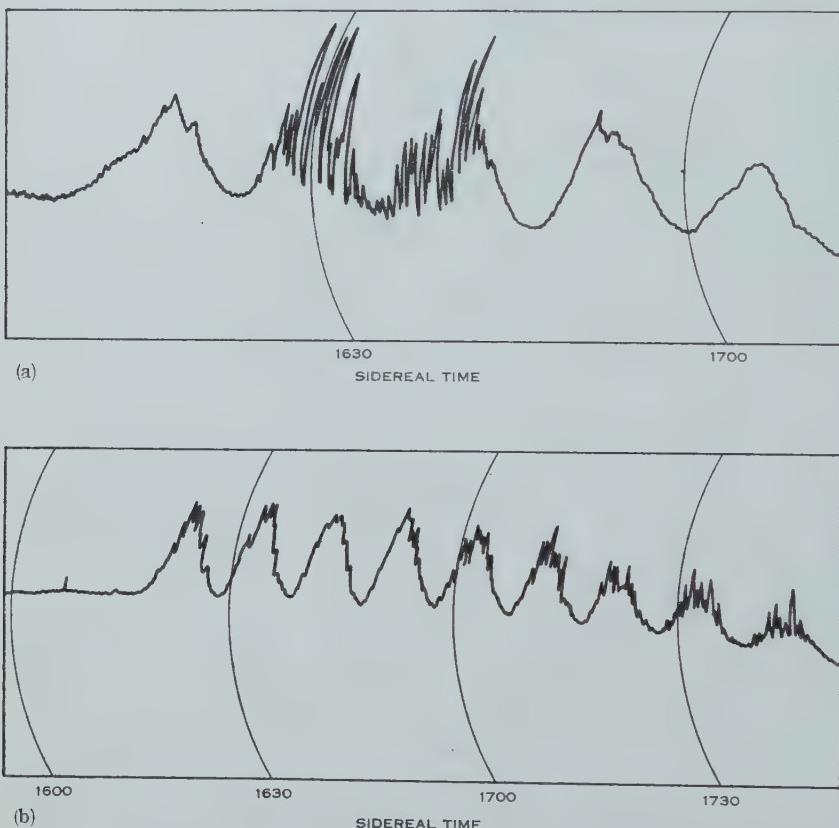
Fig. 1.—Typical records of the scintillations on interference patterns of the source in Cygnus at rising. (*a*) May 19, 1948 ; (*b*) June 11, 1949.

somewhat doubtful and these figures are shown in brackets.   The values for the three northern sources* are plotted against the month and the local time of observation in Figures 2 (*b*) and (*c*) ; the heavy lines in each figure are the means of the other three curves.   For the curves on a monthly basis, the local times of observation are spaced at intervals of approximately one-third of a day

* The declinations and mean azimuths of the four sources during the periods of observation are :   Cygnus, 40° and 145° ;   Virgo, 12° and 110° ;   Taurus, 22° and 125° ;   and Centaurus, −43° and 40°.

(actually 0·28, 0·38, and 0·34 days). With such intervals, if the scintillation index contained only a diurnal component, this would appear in the mean of the curves of Figure 2 (*b*) greatly reduced in amplitude (by a factor of five or more). The amplitude of the mean curve of Figure 2 (*b*) is, however, comparable with those of the original curves and this indicates a definite seasonal
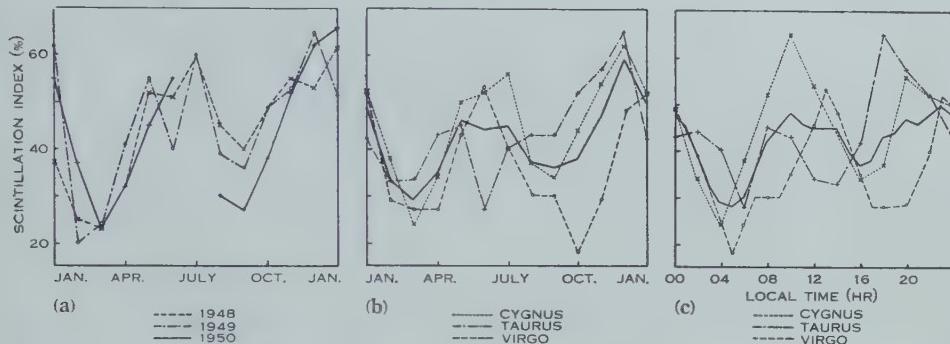
Fig. 2 (*a*).—Apparent seasonal variation in the mean monthly scintillation index for Cygnus for the 3 years 1948, 1949, and 1950.

Fig. 2 (*b*).—Mean monthly scintillation indices for Cygnus, Taurus, and Virgo plotted against the month of observation. Heavy line curve is the average of the other three and indicates the seasonal component in the variation of the indices.

Fig. 2 (*c*).—As in (*b*) but plotted against the local time of observation. Heavy line curve indicates the diurnal component in the variation of the indices.

TABLE 1

SCINTILLATION INDEX MONTHLY MEANS

Month	Cygnus-A		Virgo-A		Taurus-A		Centaurus-A	
	Number of Observations	Index	Number of Observations	Index	Number of Observations	Index	Number of Observations	Index
Jan. ..	45	52	17	52	10	42	7	54
Feb. ..	25	38	16	29	23	33	6	57
Mar. ..	57	24	28	27	35	34	16	37
Apr. ..	54	34	28	27	28	43	21	36
May ..	80	50	46	43	19	45	24	36
June ..	67	52	27	53	(3)	(27)	18	46
July ..	39	56	35	40	24	40	21	27
Aug. ..	50	37	21	30	7	43	—	—
Sept. ..	67	34	5	30	(3)	(43)	—	—
Oct. ..	66	44	15	18	(4)	(52)	(3)	(60)
Nov. ..	50	54	27	29	9	57	11	66
Dec. ..	49	65	22	48	12	65	6	73

component. Similarly, the mean curve of Figure 2 (*c*) indicates a true diurnal effect, as the effect of the seasonal component is almost eliminated by the displacement of it in the three original curves by intervals of one-third and two-thirds of a year. The two components have about the same amplitudes ;

the seasonal component has maxima about midsummer and midwinter with minima near the equinoxes, and the diurnal component has maxima near midday and midnight and minima near dawn and sunset.*

It can be seen from Table 1 that the monthly values of the scintillation index are different for Virgo and Centaurus, although the two sources are observed at about the same local time. This may possibly be due to the ray tracks of the two sources traversing the region of the atmosphere responsible for the scintillations at widely different latitudes. The ray paths for the sources through the E and F layers of the ionosphere are shown in Figure 9. The ray path for Centaurus is far to the south of those of the other three sources. It is not possible to suggest whether the seasonal or diurnal component of the scintillation index is more affected by latitude, as the Centaurus data are incomplete for several months of the year.

(c) Scintillations of a Number of Sources on the Same Day

As might be expected from the seasonal component in the scintillation index, there is fair correlation between the scintillations of sources on the same day. This has been observed in " round the clock " observations on 10 or more sources. With few exceptions all sources show some or no scintillation activity during the same period, and active or inactive periods may last for several days. The condition of the atmosphere responsible for the scintillations must therefore be widespread and must persist for several days at a time.

(d) Variation of the Scintillation Index with Frequency

In 1950 a series of observations were made of the Cygnus source on frequencies between 40 and 400 Mc/s. Each record was assessed for a scintillation index and the ratio of that index to the index for the 100 Mc/s record of the same day determined. This procedure was necessary as only a few observations were made at each frequency and the observations at the various frequencies were made at different times of the day and year. Difficulties due to seasonal and diurnal variations and day-to-day changes are thus removed by relating the observations to those on the base frequency. The complete series of results is shown in Figure 3 (a). Owing to the difficulty of assessing some of the higher frequency records where there is a low signal-to-noise ratio, the increase in the ratio beyond 200 Mc/s is questionable. More reliance can be placed on the curve of Figure 3 (b) where only records on active days, when the index on both frequencies exceeded 30 per cent., have been used.†

---

* This result is different from that of Ryle and Hewish (1950) or Hewish (1952) from observations at much higher angles of incidence in the northern hemisphere. They find a principal diurnal component in the variation of the scintillation index with a single maximum at about 01 hr. There is a slight change in the shape of the diurnal variation during the year and it has the lowest amplitude near the equinoxes.

† Hewish (1952), from observations over a 2 : 1 range of frequencies, suggests that the index is approximately proportional to the square of the wavelength. This is in fair agreement with the low frequency section of Figure 3 (b).

(e) *Variation in the Scintillation Index with the Altitude of the Source*

The scintillation index, on the average, shows a marked decline with increasing altitude of the source. This change is difficult to demonstrate from a single record or even a short series of records, as the scintillation index sometimes changes appreciably in a period of minutes. On our records the change in altitude is divided into intervals of 1° by the effect of the interference fringes. In order to reduce the labour involved in analysis we have examined only
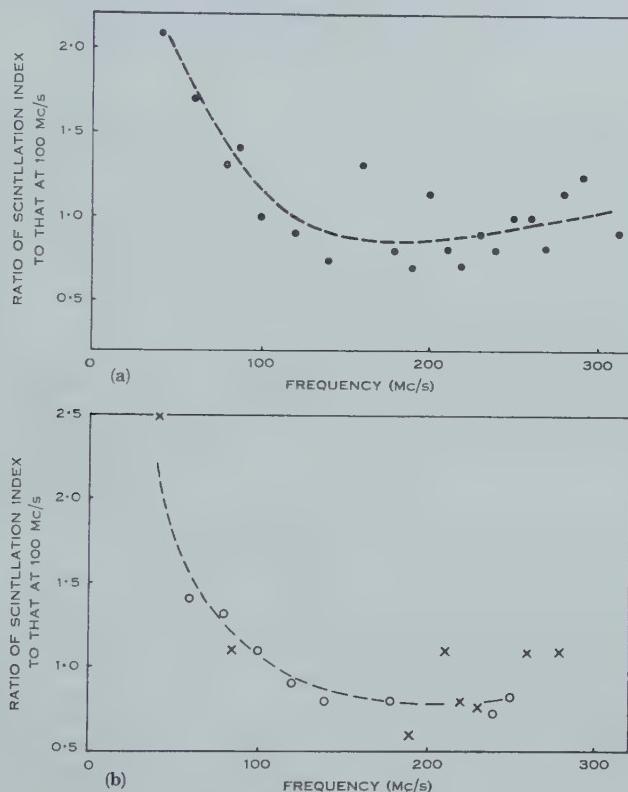
Fig. 3.—Curves showing the variation of the Cygnus scintillation index with frequency. (*a*) Includes all available records ; (*b*) only records on which the scintillation index at both frequencies was greater than 30 per cent.

Circles are means of more than and crosses means of less than four readings.

the distribution of fringes in which the scintillations are particularly severe. The fringes (which for short we shall call X-fringes) are those for which the index is almost 200 per cent. Histograms of the distribution of X-fringes against altitude for the four sources are shown in Figure 4. They show a marked decline in the number of X-fringes with altitude and reflect the general decrease in scintillation activity with the altitude of the source. These histograms should possibly tail off a little more slowly, as the eighth and ninth fringes are missing

on some records and are difficult to judge on others.   It is stressed that the
histograms do not arise from a number of records with X-fringes from 0 to 10°,
a number with X-fringes from 0 to 9°, and so on, but merely indicate that the
probability of observing an X-fringe decreases with increasing altitude.   The
number of observations of Cygnus for which the first fringe is an X-fringe (140)
is quite a small fraction of the total of observations (about 800), but the
histograms are considered to be a reliable indication of the behaviour of the
scintillations with altitude.   The same data of Figure 4 for Cygnus are presented
in Figure 5 (a) on a monthly basis ; they have been normalized by multiplying
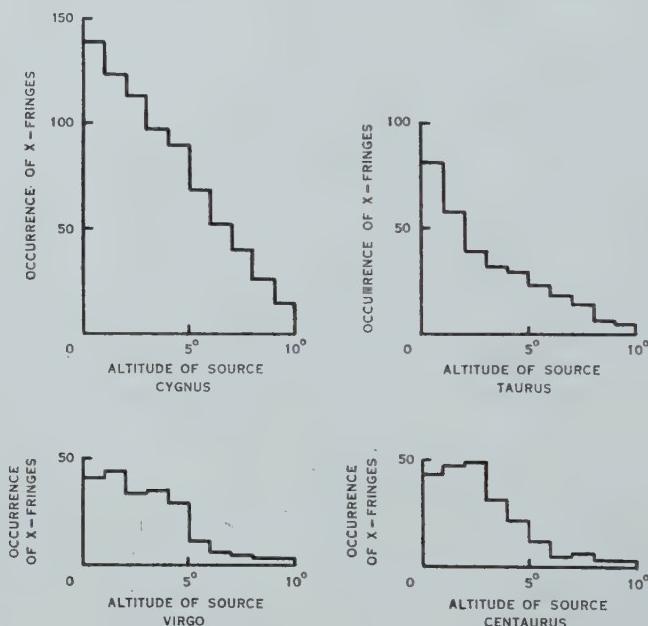
Fig. 4.—Histograms showing the number of occasions on which the
scintillation index reached 200 per cent. for successive intervals of 1°
altitude (X-fringes).   The observations refer to four sources after
their rising.   The histograms reflect the decrease in scintillation
activity with increasing altitude.

the number of X-fringes by the ratio of the average number of observations in
each month to the actual number of observations in that month.   As a general
rule the scintillations decrease rapidly with increasing altitude.   However, this
decrease is less rapid during the winter months.

It can be seen from Figure 5 (a) that the incidence of X-fringes reflects
the apparent seasonal variation in the scintillation index.

(f) Scintillation Rates for Four Sources

To determine the mean scintillation rate for each source a group of scintilla-
tions in an interval of 5–10 min was counted on each record.   The place chosen
corresponded to an interval of between 2 and 4° in the altitude of the source.

This ensured that the scintillations were of sufficiently large amplitude to make an accurate count and yet avoided peculiar effects that sometimes occur near grazing incidence. The data obtained are presented in Table 2 and Figure 6, the latter showing the distribution of scintillation rates among the days of observation.

In the case of Cygnus, sufficient data were available to study the variation of scintillation rate throughout the year. The results are presented as histograms in Figure 5 (b), along with the histograms showing the variation in the incidence of scintillations with altitude. The scintillation rate increases appreciably in May and June, the 4-year averages being 1·28 and 1·40 against the yearly mean of 1·15 per minute. This feature has shown up even in the individual years, the monthly values for June and October being 1·40 and 1·14 for 1948, 1·34 and 1·12 for 1949, and 1·60 and 1·06 for 1950 from about 10 observations in each month.
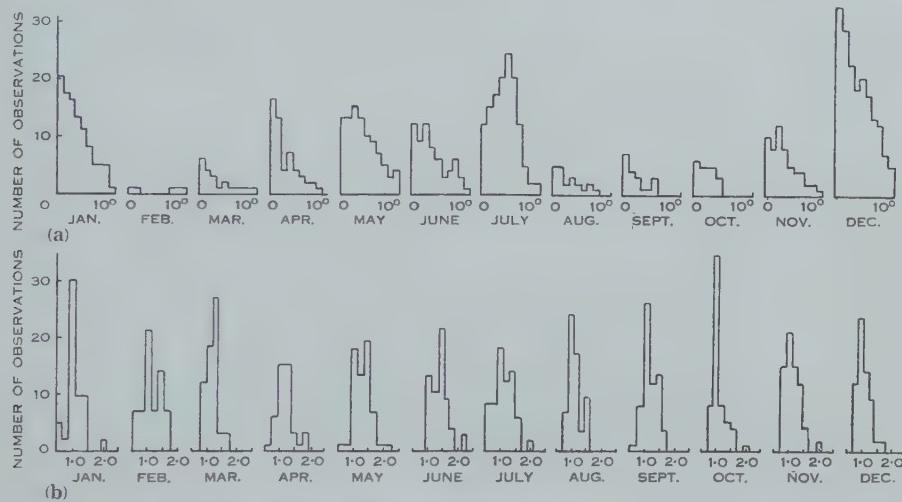
Fig. 5 (a).—Incidence of scintillations with altitude for Cygnus. The data of Figure 4 on a monthly basis. Note the change in distribution of the X-fringes with altitude in the winter months, indicating a slower decline in the scintillation activity with altitude.

Fig. 5 (b).—Seasonal variation of scintillation rate for Cygnus. The data of Figure 6 on a monthly basis. Note the higher average scintillation rates in May and June.

(g) Variation of the Scintillation Rate with Altitude

It was not possible to make any accurate estimate of the change of scintillation rate with altitude. A considerable section of each day's record was used in the foregoing analysis. In a similar section of the records at an appreciably greater altitude (say 7°) the smaller amplitude of the scintillations does not permit accurate counting. Visual inspection of the records does not suggest any marked change in the rate between altitudes of 1 and 10°.

(h) Variation of the Scintillation Rate with Frequency

The same records as were used in the analysis of Section II (d) were examined to find out whether there was any change in the scintillation rate with frequency.

F

Here again, owing to day-to-day variations in the rate and the small number of observations at each frequency, it was necessary to relate the scintillation rate at a particular frequency to that of the base frequency of 100 Mc/s on the same day. The results for six observations at each of five frequencies are presented
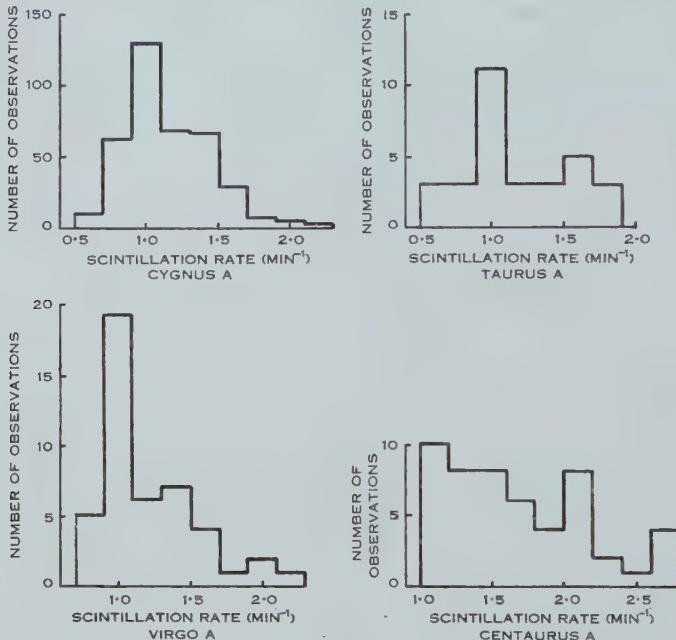
Fig. 6.—Histograms showing the distribution of scintillation rates for four sources. Each observation refers to an individual night and is the average rate during a period of 5–10 min when the altitude of the source is between 2 and 4°.

TABLE 2

MEAN SCINTILLATION RATES OF FOUR SOURCES

Source	Mean Scintillation Rate (min^{-1})	No. of Records Used	Approx. No. of Scintillations Counted
Cygnus	1·14	390	2400
Virgo	1·18	45	270
Taurus	1·24	38	200
Centaurus	1·80	51	400

in Table 3 where the differences are not greater than the experimental uncertainty. It is believed that the scintillation rate is independent of frequency over this range.* Owing to the interference method of observation and the rapid change of the scintillation index with frequency, it is generally difficult to study the

* A similar result is reported by Hewish (1952) over a 2 : 1 frequency range.

TABLE 3

VARIATION OF SCINTILLATION RATE WITH FREQUENCY

Frequency (Mc/s)	60	100	120	140	180	240
Ratio of scintillation rate to that at 100 Mc/s	0·89	1·00	0·91	0·96	0·92	1·02

correlation of individual scintillations over a range of frequencies. However, when this is possible, some similarity can be seen between groups of scintillations over a range of two to one in frequency.

(i) *Relation between the Scintillation Rate and Index*

Owing to the slow recorder speed of 6 in/hr used in most of the observations, it is difficult to study the relation between the scintillation rate and index. However, a small number of observations were made in New Zealand in 1948
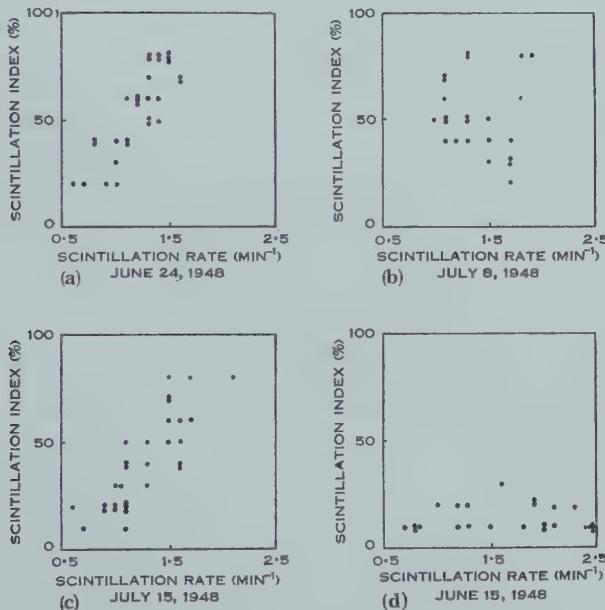
Fig. 7.—Correlation diagrams of the scintillation indices and rates for 4 nights' observation of Cygnus. Each point represents an interval of 4 min.

using a much higher recorder speed.* On most nights the rate and index showed large independent variations or one quantity remained constant while the other varied, but on about one-third of the nights there appeared to be some

* From the limited number of observations no change was found in the average scintillation rate with altitude. On occasions violent scintillations were observed up to altitudes of 14°. This is in agreement with the Sydney observations for the winter months (Section II (e) and Fig. 5 (a)).

association between the variations in the two quantities. The results for four of the nights are shown in Figure 7, each dot representing the mean value for a period of about 4 min.

Scintillations of the type represented by Figures 7 (a) and (c) have been occasionally observed in groups lasting 10–30 min on otherwise inactive days.

SIDEREAL TIME
CYGNUS MARCH 21, 1950

SIDEREAL TIME
CYGNUS MARCH 20, 1950
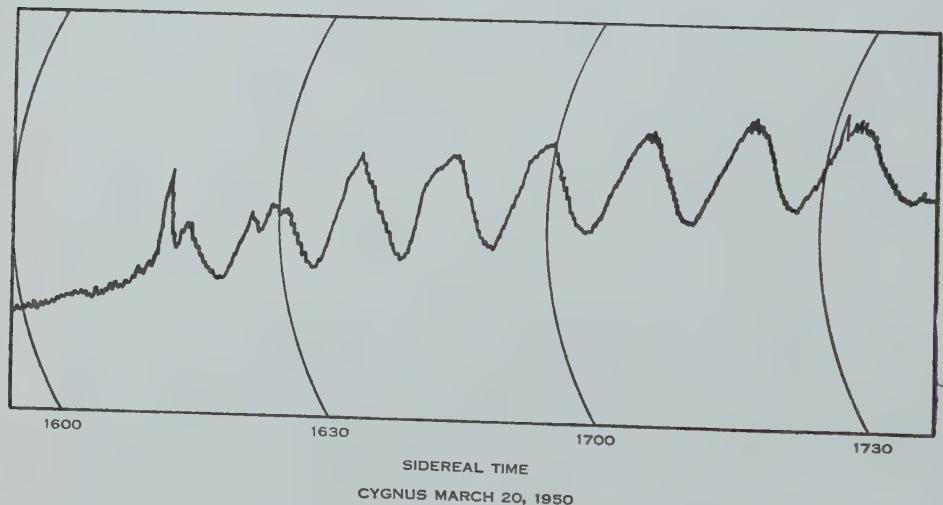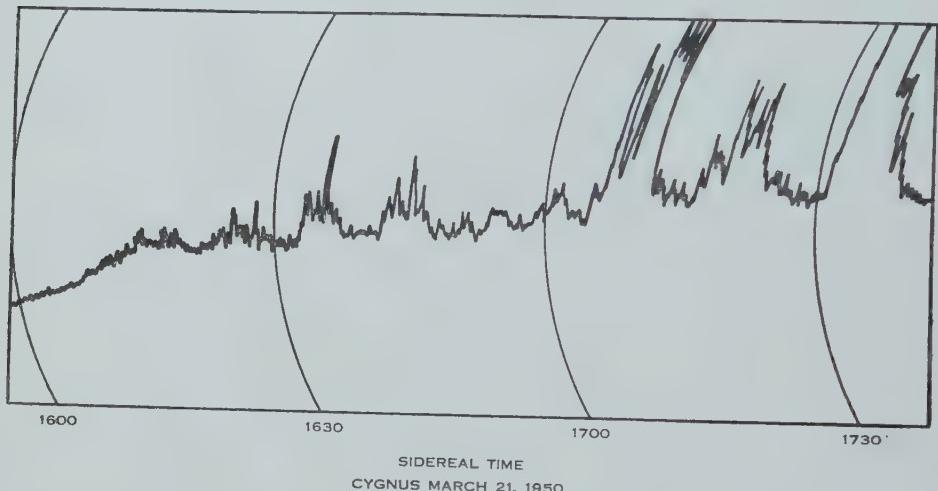
Fig. 8.—The upper record shows strong absorption and a very high scintillation rate for Cygnus at rising. The lower record, obtained the previous day, is typical for March.

The activity commences with low amplitude scintillations of long duration, and builds up to high amplitude scintillations of short duration. The reverse occurs as the activity declines. It may be that the mechanism responsible for the scintillations is different from the usual one in these cases.

(j) Other Phenomena

No association has been observed between the scintillation activity and the atmospheric conditions at ground level, so that the troposphere is unlikely to be the source of the scintillations.

No obvious long- or short-term connexion has been found between the scintillations and solar activity ; no change is observed in the scintillations at sunrise ; the long series of Cygnus observations has been examined for periods of the order of 27 days without results and the average scintillation activity between 1948 and 1950 (cf. Fig. 2) has remained practically constant while the sunspot numbers declined appreciably. The average sunspot numbers for the three years were 135, 135, and 85.

Possible exceptions to the foregoing are three unusual records out of a total of about 2000. One of these is illustrated in Figure 8. The lower record is typical for Cygnus in March, rising just before dawn. The unusual one (upper record) shows strong absorption of the signal and an abnormally high scintillation rate. Some of the scintillations are too fast for accurate reproduction. On two of the three occasions when such records were obtained, a visible aurora* was reported from southern Australia at about the time of observation. On the third occasion no aurora was reported but high level radio emission was received from the Sun on the preceding day.

III. Scintillations and their Association with Sporadic E

(a) Summary of Experimental Evidence

Before discussing the scintillations and their association with ionospheric phenomena the main points of the experimental evidence may be summarized as follows.

The scintillation index
(1) Shows both seasonal and diurnal variations.
(2) Decreases with decreasing wavelength.
(3) Decreases rapidly with increasing altitude. For Cygnus this decrease is less rapid in the winter months.

The scintillation rate
(1) Is independent of wavelength.
(2) Is different for sources of different declination. For Cygnus the rate is higher during the winter months.

(b) Correlation with Ionospheric Phenomena

Some difficulty is inevitably encountered in attempting to correlate the low altitude scintillations with local ionospheric phenomena. This is illustrated in Figure 9 which shows the ray tracks of the four sources in the E and F layers of the ionosphere during the periods of observation in relation to the ionospheric recording stations. There are considerable distances between the ray tracks and the nearest stations but the ionospheric condition responsible for the scintillations might, however, be expected to be fairly widespread (Section II (e)). We have

---

* A similar association of high scintillation rate and auroral phenomena is reported by Little and Maxwell (1951) and with magnetic disturbances by Hewish (1952).

therefore sought correlation between the scintillations of the three northern sources and ionospheric data from Brisbane and between scintillations of Centaurus and data from Hobart and Macquarie Island. We have used the hourly values given in the Ionospheric Predictions, Series D, of the Ionospheric Prediction Service of the Commonwealth Observatory. In assessing any indices the hourly value nearest to the time of observation and the two values on either side have been considered.

(c) Correlation with Spread F

Ryle and Hewish (1950), Little and Maxwell (1951), and Mills and Thomas (1951) have suggested that the scintillations are associated with disturbances in the F layer, in particular with the condition known as spread F. Such an association with the low altitude scintillations appears unlikely as the incidence
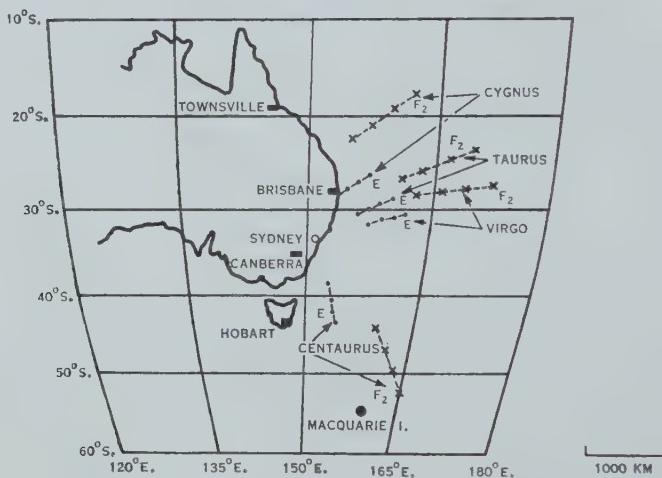
Fig. 9.—Sketch showing sections of the ray tracks of the four sources through the E and F_2 layers (at 100 and 400 km) of the ionosphere. These sections refer to the maximum period of observation of the sources each day and correspond to a range of altitude of 0 to 12°; the dots and crosses are at intervals of 4° altitude. The positions of the ionospheric recording stations on the east Australian coast are also shown.

of spread F increases rapidly with geographic or magnetic latitude. It can be seen from Figure 9 that the ray track of the Centaurus source lies between magnetic latitudes of 50 and 60 °S., whereas the ray tracks of the northern sources are near latitude 20 °S. in the F_2 layer, yet the mean scintillation indices for the whole year are practically the same for all the sources. In spite of this, correlation was sought between the Centaurus scintillations and spread F at Macquarie Island and Hobart but with no result. A negative result was also obtained between the scintillations of Cygnus and spread F at Brisbane. However, when the analysis was restricted to the three winter months a small correlation was found. (Mills and Thomas's observations were mainly concerned with winter months.)

(d) Correlation with Sporadic E

The seasonal and diurnal components in the scintillation index suggest an association with sporadic E (see McNicol and Gipps 1951). A day-to-day correlation of the two phenomena was examined on the following basis. The daily scintillation index was divided into three classes :

H, high : index 60 per cent. or greater.
M, medium : index between 30 and 60 per cent.
L, low : index 30 per cent. or less.

Similar classifications were assigned to the sporadic E recorded at Brisbane (or Hobart in the case of Centaurus) :

H, high : critical frequency high for the month concerned.
M, medium : echoes present only part of the time, or critical frequency low for the month concerned, or both.
L, no echoes observed.

TABLE 4

CORRELATION BETWEEN CYGNUS SCINTILLATION INDEX AND SPORADIC E AT BRISBANE

Sporadic E at Brisbane	Cygnus Index		
	High (H)	Medium (M)	Low (L)
High (H)	57	18	28
Medium (M) ..	35	31	41
Low (L)	27	34	92

The results of the analysis of the Cygnus and Brisbane data for the years 1949 and 1950 are shown in Table 4. The number of correlations is 180 or 50 per cent. of the total, the number of partial correlations is 128 or 35 per cent. of the total, and the number of anti-correlations 55 or 15 per cent. of the total. The percentages that would result from pure chance are 34, 40, and 26 respectively. If we restrict the analysis to the very definite cases of high or low indices, thus neglecting cases in which the sporadic E is probably patchy, we obtain the results in Table 5 for the four sources. In this table the figures that would result from pure chance are shown in brackets. In each case there are more than 20 per cent. more correlations than would be expected on the basis of pure chance.* The probabilities against the observed results occurring by chance are less than 1 in 15,000 for Cygnus, 1 in 360 for Virgo, 1 in 25 for Taurus, and 1 in 12 for Centaurus, the lower values being due to the smaller number of observations. The high significance of the number of correlations indicates a strong connexion between the low altitude scintillations and sporadic E.

* These figures are the same (or better if partial correlations are included) as those obtained between scintillations at high angles of incidence and spread F by observers in the northern hemisphere (Ryle and Hewish 1950 ; Little and Maxwell 1951). The data for Cygnus were examined in periods of 3 months but the percentage of correlations was not found to vary with the time of year.

(e) *The Scintillation Index and the Critical Frequency of Sporadic* E

A comparison between the mean monthly scintillation index for Cygnus and two indices describing sporadic *E* at the time of observation of the source is shown in Figure 10. These two indices are the percentage of days in each month when sporadic *E* was observed and the mean critical frequency of the sporadic echoes on the days observed. The results are averages for the years 1949 and 1950. It will be seen that, although the form of the three curves is similar, the critical frequency of the echoes is lower in midwinter than midsummer, and the two maxima of the scintillation index curve are about equal. The reason for the difference in the relative heights of the maxima is probably to be found in the seasonal characteristics of the sporadic *E*. Sporadic *E* at Brisbane has been studied extensively by McNicol and Gipps (1951). They

TABLE 5

CORRELATION BETWEEN SCINTILLATION INDICES OF FOUR SOURCES AND SPORADIC *E* AT NEAREST IONOSPHERIC STATIONS

Sporadic *E* at Brisbane	Cygnus Index		Sporadic *E* at Brisbane	Virgo Index	
	H	L		H	L
H	57	28	H	27	22
	(34)	(50)		(16)	(33)
L	27	92	L	9	53
	(50)	(70)		(20)	(42)

Sporadic *E* at Brisbane	Taurus Index		Sporadic *E* at Hobart	Centaurus Index	
	H	L		H	L
H	18	7	H	8	6
	(12)	(13)		(5)	(9)
L	7	20	L	2	13
	(13)	(14)		(5)	(10)

give an overall picture of the incidence of sporadic *E* as showing both seasonal and diurnal components, the former with minima near the equinoxes and the latter with minima near dawn and sunset, that is, the same as is found for the scintillations of the northern sources. They find two distinct types of sporadic *E* : one, predominant in summer, has high critical frequencies and strongly blankets the upper layers of the ionosphere, and the other, predominant in winter, has lower critical frequencies but does not blanket the upper layers. The summer type is present most of the year but only occurs on about one day in four in winter and for shorter periods. As there is no blanketing by the winter type it must have pronounced lateral irregularities. It is suggested that if one of the factors involved in the production of scintillations is the relative retardation of two adjacent rays through the E_s layer, the winter type with its lower electron concentration and more pronounced lateral irregularity might be just as effective as the summer type with a more uniform but higher electron concentration.

It is interesting to note that McNicol and Gipps suggest a possible correlation between the incidence of spread echoes from the E_s and F layers during the winter nights. This may explain the slight correlation between the incidence of Cygnus scintillations and spread F at Brisbane found for winter nights (Section III (c)).

(f) The Size of the Irregularities in the E Layer

Other writers have provided fairly satisfactory explanations of the scintillation phenomena in terms of a diffraction screen formed by irregularities in the relevant layer of the ionosphere. Hewish (1952) has shown that the phase deviation of the wave emergent from the screen governs both the amplitude and scale of the diffraction pattern on the ground. Determination of the size of the
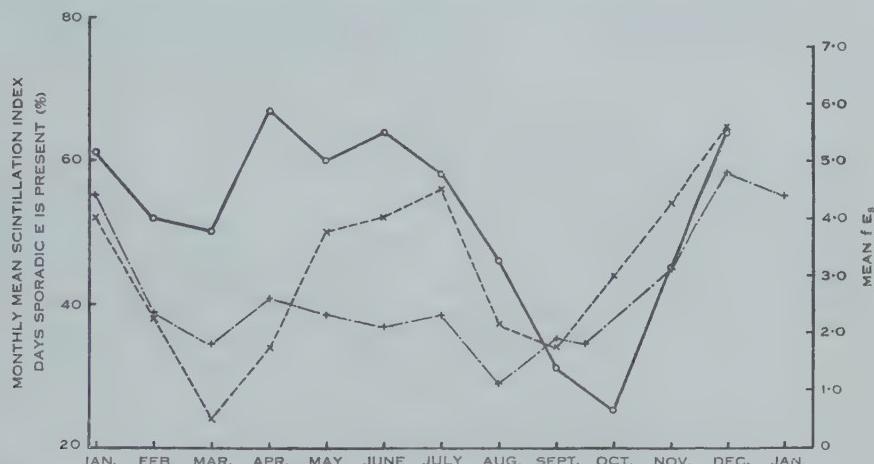
Fig. 10.—The annual variation in the mean monthly scintillation index for Cygnus ($\times$————$\times$), the percentage of days during the month on which sporadic E is observed at the time of the Cygnus observation (o————o) and the mean critical frequency ($+$——·——$+$) on those days. Ionospheric datum is for Brisbane and all data are averages for the years 1949 and 1950.

irregularities in the screen requires in general a knowledge of the lateral extent of the phase and amplitude deviations on the ground. For phase deviations at the screen of less than 1 radn the scale of the diffraction pattern on the ground is the same as that of the irregularities in the screen, but for deviations greater than 1 radn the scale of the ground pattern is smaller. In the latter case the scale decreases as the wavelength increases.*

The scintillations result from the relative motion of the ray track of the source and the diffraction screen. In our case, as the scintillation rate is independent of wavelength between wavelengths of 1 and 5 m, we can assume

---

* Little (1951) suggests that the diffraction patterns on the ground are similar for a range of wavelengths if at the longest wavelength the area over which there is correlation on the ground is larger than the size of the first Fresnel zone at the screen. In such a case the scale of the pattern on the ground is the same as that of the irregularity of the screen.

that the phase deviation at the screen is less than 1 radn over this range.   Hence
if the diffraction theory is applicable to our case the apparent size of the
irregularities can be deduced from the scintillation rate and the velocity of the
ray track through the E layer (assumed stationary).   The relevant data are
given in Table 6.   The fifth column, obtained by dividing the velocity of the
ray track by the scintillation rate, gives the apparent size of the irregularities.
These sizes are much greater than those found by other ionospheric measurements
and are more of the order of individual clouds of sporadic E.

It is reasonable to assume that over the relatively small distances between
the ray tracks of the three northern sources (of Fig. 9) the true sizes of the
irregularities would not change by the amounts indicated in Table 6.   The
differences in the apparent sizes can be explained by postulating irregularities
which are elongated in an east-west direction, or by taking ionospheric winds
into account, or both.   Munro (1950) finds that near Sydney winds of the order
of 3 to 4 km/min exist in the E layer ; their directions are towards the north-west

TABLE 6

RAY TRACKS OF SOURCES AND SCINTILLATION RATES

| Source | Velocity of Ray Track | | Scintillation Rate (min⁻¹) | $\dfrac{\text{Velocity}}{\text{Rate}}$ (km) (*Apparent* Size of Irregularities) |
	Speed (km/min)	Direction of Travel of Ray Track		
Cygnus	4·1	30° S. of W.	1·14	3·6
Taurus	7·2	20° S. of W.	1·24	5·8
Virgo	8·0	10° S. of W.	1·18	6·8
Centaurus	6·1	75° N. of W.	1·80	3·3

in summer but more towards the north in winter.   A wind of 3·5 km/min
towards the north-west would reduce the discrepancies of Table 6, giving the
sizes as 3·8 km for Cygnus, 4·9 km for Taurus, and 5·3 km for Virgo.   In
addition, the change of wind direction towards the north in winter would explain
the higher scintillation rate observed at that time for Cygnus.   However, the
discrepancy between Centaurus and the other three would be further increased
by such winds but it is likely that both the size of the irregularities and the
winds are different in the part of the E region crossed by the ray track of
Centaurus.

If the irregularities were in the F layer and Munro's wind characteristics
for this layer adopted, the size of the irregularities would average 17 km for
the three northern sources, that is, much greater than the sizes indicated by
spaced receiver observations.

(g) *The Decrease of Scintillations with Altitude*

The observed decrease in the scintillation index with altitude is much
greater than would be expected merely on the basis of the change in the path
length through the E layer.   A similar result has been obtained by Little and

Maxwell (1951) and Hewish (1952). On the diffraction theory they suggest that up to a certain distance the index would increase with the distance from the screen. In our case the change in distance is $2:1$ during the period of observation.

IV. References

BOLTON, J. G., and STANLEY, G. J. (1947).—*Nature* **161** : 312.

HEWISH, A. (1952).—*Proc. Roy. Soc.* A **214** : 494.

HEY, J. S., PARSONS, S. J., and PHILLIPS, J. W. (1946).—*Proc. Roy. Soc.* A **192** : 425-45.

LITTLE, C. G. (1951).—*Mon. Not. R. Astr. Soc.* **111** : 289.

LITTLE, C. G., and MAXWELL, A. B. (1951).—*Phil. Mag.* **42** : 267.

LOVELL, A. C. B., and LITTLE, C. G. (1950).—*Nature* **165** : 423.

McNICOL, R. W. E., and GIPPS, G. DE V. (1951).—*J. Geophys. Res.* **56** : 17-31.

MILLS, B. Y., and THOMAS, A. B. (1951).—*Aust. J. Sci. Res.* A **4** : 158-71.

MUNRO, G. H. (1950).—*Proc. Roy. Soc.* A **202** : 208-23.

RYLE, M., and HEWISH, L. A. (1950).—*Mon. Not. R. Astr. Soc.* **110** : 381-94.

SMITH, F. G. (1950).—*Nature* **165** : 422.

THE RADIO BRIGHTNESS DISTRIBUTIONS OVER FOUR
DISCRETE SOURCES OF COSMIC NOISE

By B. Y. Mills*

Summary

Brightness distributions have been obtained across the four radio sources, Cygnus–A, Taurus–A, Virgo–A, and Centaurus–A, using a two-aerial interferometer of a special type in which the aerial spacing and the azimuth angle of the axis may be varied over a wide range. Radio isophotes have been constructed for three of the sources from these results, making some simple assumptions as to their form. The isophotes bear some relation to optical features of the nebulae with which the sources have been identified and their radio and optical sizes are similar. The remaining source, Cygnus–A, is unfavourably situated for observations from Sydney, and the results are less complete.

I. Introduction

The nature of the discrete radio sources has been the subject of speculation since their discovery by Bolton and Stanley (1948). None of the early measurements gave any hint of their angular sizes which could well have been comparable with those of stars. However, some measurements of the positions of three of the stronger sources, Taurus–A, Virgo–A, and Centaurus–A, carried out by Bolton, Stanley, and Slee (1949) led them to suggest identifications of these sources with nebulae of unusual types. Later Mills and Thomas (1951) obtained a position for Cygnus–A which was very close to that of another faint nebula, although at that time no abnormality in the nebula was suspected.

The existence of localized sources of radiation with angular extents of the order of a degree was then demonstrated by Mills (1952a) and Piddington and Minnett (1952), some of whose preliminary results were described by Bracewell (1952). Although only one of these sources, Centaurus–A, could definitely be associated with a nebula, the large angular sizes of the others indicated a similar origin.

Further accurate position measurements (Smith 1951 ; Mills 1952b) appeared to confirm the identifications of the Taurus, Virgo, and Centaurus sources suggested earlier, and the position of the Cygnus source was found to be practically coincident with that of the faint nebula. Baade and Minkowski (1953a) found that this nebula consisted of two galaxies in collision and that the interstellar gas had a very high degree of excitation. They also found a peculiar and very faint galactic nebula in the position obtained by Smith for the Cassiopeia source.

At this stage it was clear that final evidence for the identifications would be agreement between the dimensions of the radio sources and the nebulae. Simultaneous attempts were made in Australia and England to construct

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

equipment with the necessary resolving power, of the order of 1′ of arc, to measure the radio sizes. Because of the high resolving power required, pencil-beam aerials were impractical and consequently interferometric methods were employed. These attempts were successful at approximately the same time and preliminary results have now been described. In England the sources Cassiopeia and Cygnus–A were observed by Brown, Jennison, and das Gupta (1952) and Smith (1952a, 1952b), and in Australia the sources Cygnus–A, Taurus–A, Virgo–A, and Centaurus–A,* by Mills (1952c). All five sources were resolved and found to have angular sizes comparable with their associated nebulae, thus verifying the identifications, but the measurements gave insufficient information from which to deduce the shapes and brightness temperatures of the sources.

The present paper gives the results of much more extensive observations carried out on the same four sources. They have led to more detailed knowledge. The equipment and techniques employed are also described more fully than in the preliminary note. It is found that there are certain similarities between the radio and optical features of the nebulae.

II. THEORY

The fundamental method of obtaining the brightness distribution across a source of radiation by means of interferometer observations was pointed out by McCready, Pawsey, and Payne-Scott (1947). By taking a number of observations with different aerial spacings along a common axis, it is possible to derive the " integrated " brightness distribution over the source, where the integrated brightness is the integral of the emission along a line at right angles to the projection of the interferometer axis on the source. If the radio isophotes are known to be circularly symmetric, this series of observations is sufficient to determine the actual surface brightness distribution. If the isophotes are central ellipses, it will be shown later that three sets of observations with three different interferometer axes will determine the actual brightness distribution. For more complex shapes even more sets of observations are necessary. In this work three sets of measurements have been made, extensive measurements with an interferometer axis at one azimuth angle (east-west) and one or two measurements at different azimuths. While there is no certainty that a true " radio picture " of a source can be constructed in this way, it is probable that valuable information can be obtained by comparing the radio and optical pictures. If very much more detailed information is required, pencil-beam methods are likely to lead to solutions more efficiently.

The relation between the aerial spacing, the distribution across the source, and the amplitude and phase of the interference pattern is derived in Appendix I. It is given by

$$I(\theta) = \int_0^\infty A_n \cos (\varphi_n + 2\pi n \ \theta) \mathrm{d}n, \ \ldots\ldots\ldots\ldots (1)$$

* These sources were designated 19+4, 05+2, 12+1, and 13—4 respectively by Mills (1952a, 1952b).

where $I(\theta)$ is the " integrated " brightness distribution across the source, a function of angle θ, n is the effective aerial spacing measured in wavelengths (see below), and A_n and φ_n are the relative amplitude and phase angle of the interference pattern.

The effective aerial spacing n is the projection on the wave front of the actual aerial spacing. It is related to the actual aerial separation, n', the azimuth of the interferometer axis, A, and the zenith angle of the source, Z, by

$$n = n'(1 - \cos^2 A \, \sin^2 Z)^{\frac{1}{2}}. \quad \dots\dots\dots\dots\dots \quad (2)$$

For a source in which a sharp concentration is superimposed unsymmetrically on a larger diffuse area, the interference pattern will be the vector sum of two components due to the sharp concentration and diffuse area respectively. From the Fourier transform equation (1) it will be seen that the amplitude of the diffuse area component will fall off rapidly as the aerial spacing is increased ; the sharp concentration component will not be reduced in the same proportion. The phase difference between these two components also depends on the aerial spacing, as well as the relative positions of the two sections of the source. Adding these components together, the resulting interference pattern will have an amplitude and phase which depend on the aerial spacing. At large spacings the diffuse area component will be negligible and the apparent position of the source will be that of the sharp concentration.

Determination of the phase angle φ_n implies the measurement of apparent position at each spacing. This could be done only with great difficulty with the present interferometer, so each source is assumed to be symmetrical. This allows φ_n to be put equal to zero in equation (1) and reduces it to a Fourier cosine transform. The assumption of symmetry is not unreasonable in view of the approximate symmetry of the nebulae associated with the radio sources. The same assumption has been made by Stanier (1950) and Machin (1951) in their work on the distribution of brightness across the Sun. One effect is to concentrate unsymmetrical features of the distribution into the central regions so that if the distribution is complex it is very likely that the brightness of the centre will be increased at the expense of the outer regions.

The surface brightness isophotes of a source may be investigated from the integrated brightness distributions along various axes. If the isophotes are concentric ellipses, then three axes are needed. In the general case for such ellipses the shape as well as scale varies along different axes, and a rather complicated normalizing procedure is required. In this work, however, it is found that the observations are consistent with distributions with the same shape but different scales along different axes. So the isophotes are assumed to have the same ellipticity and orientation.

The procedure for such a model is as follows. Firstly, for each axis the corresponding integrated distribution is used to determine the radial brightness distribution $B(r)$ for a circularly symmetrical model. It can easily be shown that $B(r)$ is the solution of the Abelian equation

$$I(\theta) = 2 \int_{\theta}^{\infty} \frac{B(r)r}{\sqrt{r^2 - \theta^2}} dr. \quad \dots\dots\dots\dots\dots \quad (3)$$

For each axis for the elliptical model it can be seen that the value of $B(r)$, determined by equation (3) for any value of θ, is now the brightness of the isophote which is tangential to the line $\theta = $ constant. From the distributions of $B(r)$ corresponding to three different axes the complete isophotes may be constructed, the centre and three tangents to each elliptical isophote being known. This procedure is followed through in detail for a particular set of observations in Section V (a).

Machin (1951) has shown that $B(r)$ may be obtained directly from the relative amplitude curve by means of a Bessel transform. Here, however, the quantity $I(\theta)$ has been calculated first as it involves less assumptions and is therefore more fundamental; also it may be obtained with much greater accuracy. $B(r)$ has then been calculated from equation (3). An approximate solution to this equation has been developed which can be obtained very rapidly with an accuracy appropriate to the many assumptions involved.

The projection of the interferometer base across the source at transit will make some angle A' with the meridian. This angle is given by

$$\tan A' = \tan A / \cos Z, \quad \ldots\ldots\ldots\ldots\ldots (4)$$

where A is the interferometer azimuth and Z the zenith angle of the source.

The preceding analysis gives only the relative values of the surface brightness. Absolute values are required, however, and it is usual to quote the equivalent brightness temperature. At radio frequencies the Planck formula for thermal radiation reduces to the Rayleigh-Jeans form which, for constant frequency increments, is

$$T_0 = \frac{S\lambda^2}{2k \int b \, d\Omega}, \quad \ldots\ldots\ldots\ldots\ldots\ldots (5)$$

where T_0 is the central temperature, S is the flux density due to the source, k is Boltzmann's constant, λ is the wavelength, b is the normalized value of the surface brightness, and $d\Omega$ is an element of solid angle in steradians.

III. Equipment

Basically the equipment consists of a two-aerial interferometer, comprising a large fixed aerial and a small portable aerial, in which the radio-frequency connection between the portable aerial and the receiver is obtained by a radio link. By this means great flexibility is achieved in the possible placing of this aerial and spacings up to 10 km or more may be used.

The operating frequency of the interferometer is 101 Mc/s. It was chosen primarily because equipment of this frequency was available and it results in a convenient size for the portable aerial. A photograph of this aerial, which consists of two Yagis connected in parallel, is shown in Plate 1. The transmitting aerials for the radio link are also shown at the top of the vertical pole. The battery-operated receiving and transmitting equipment is in the trailer. The fixed 101 Mc/s receiving aerial is a broadside array with an area of 50 m². It has been described elsewhere (Mills 1952a).

A simplified block diagram of the complete equipment from which the operation can be followed is shown in Figure 1.   Between it and the simpler types of interferometers described previously there are several differences which require detailed discussion.   First is the radio link itself.   Both the 101 Mc/s signal from the radio source, which is converted in frequency before transmission in order to avoid interaction effects, and the local oscillator frequency are transmitted at a level of about ½ W.   When these two signals are mixed at the receiving end of the link, the original 101 Mc/s signal is reconstructed with its phase changed by a constant amount.   It is then available for mixing with the signal from the fixed aerial to form the interference pattern, although it is not practicable to do so immediately for several reasons.
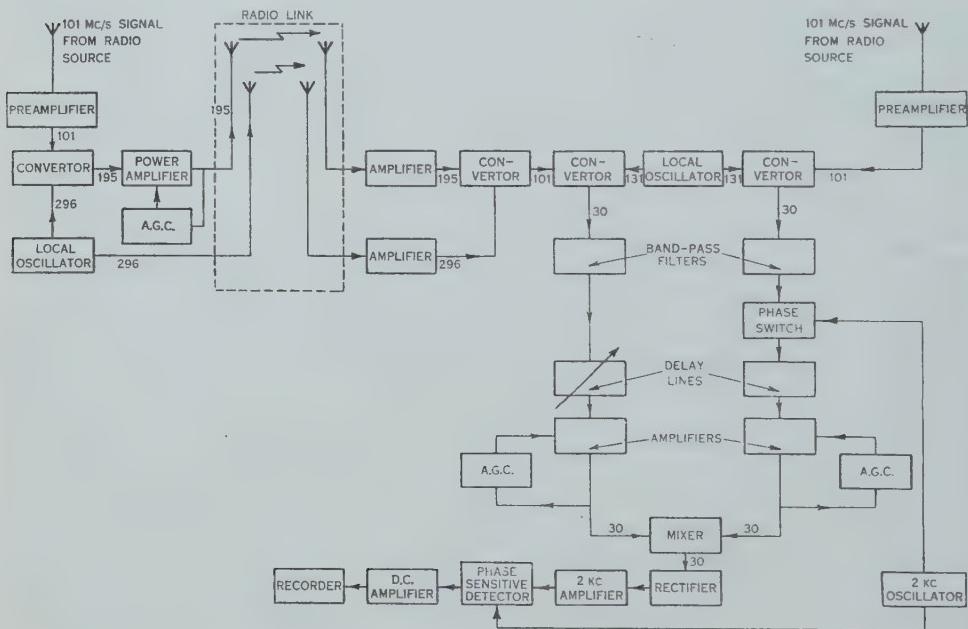
Fig. 1.—Block diagram of the complete equipment.   Operating frequencies (in Mc/s) are indicated.

First, it is necessary to control the propagation times of the signals from each 101 Mc/s aerial to the mixing point.   When these are equal and the radio source is in a plane at right angles to the interferometer axis, all the component frequencies within the pass band of the equipment arrive at the mixing point in the same phase and the interference pattern will have its maximum amplitude. When the propagation times are unequal, the phases of the component frequencies will be unequal and the pattern will be reduced in amplitude.   The effect becomes more pronounced at large aerial spacings, and with a spacing of 10 km it confines the response of the interferometer to an angle of about $1\frac{1}{2}°$ on either side of the central collimation plane.   The collimation plane may be moved by changing the propagation time between one aerial and the mixing point.   An analysis

of the effect in the case of a sea interferometer has been given by Stanley and
Slee (1950). When the interferometer axis is east-west, the propagation times
must be adjusted for equality at each different aerial spacing. When the aerial
axis is not in an east-west direction, adjustment is required for different declina-
tions in addition, since the collimation plane must be adjusted to intersect the
meridian at the zenith angle of the source under observation. The most
convenient means of obtaining the required variation is by the use of mercury-
filled acoustic delay lines. It is difficult to construct a mercury line for very
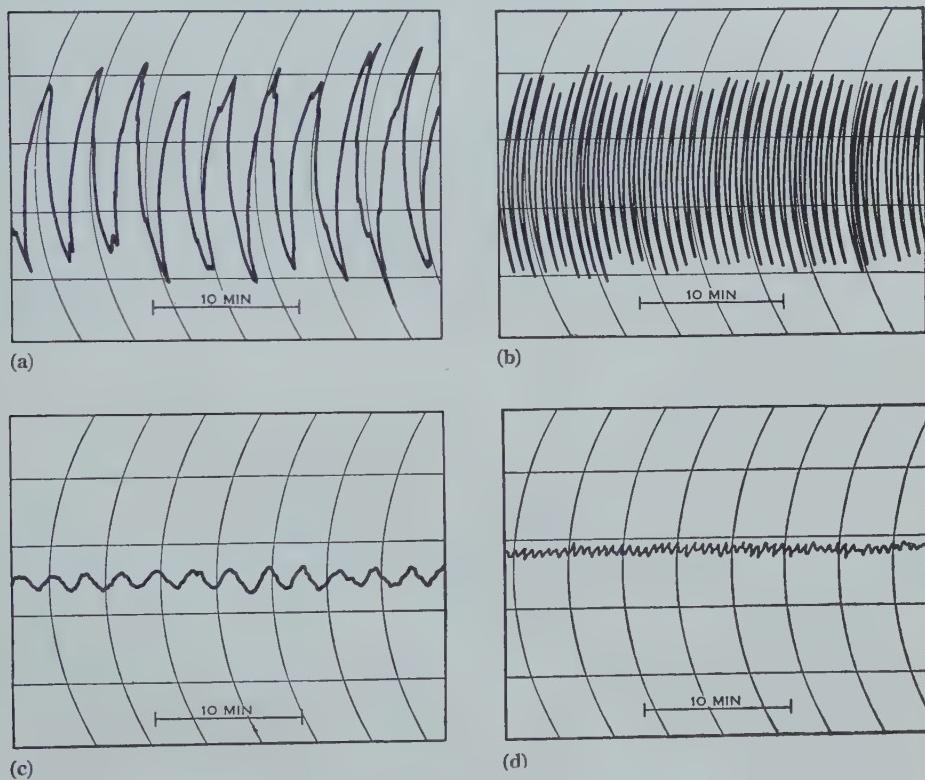
Fig. 2.—Some sample records.

(a) Cygnus–A.   Aerial spacing=0·29 km, E.-W., 29.ix.52.
(b) Cygnus–A.   Aerial spacing=1·25 km, E.-W., 29.iii.53.
(c) Virgo–A.   Aerial spacing=0·29 km, E.-W., 29.ix.52.
(d) Virgo–A.   Aerial spacing=1·25 km, E.-W., 3.iv.53.

short delays so that it is necessary to insert a delay in both signal paths. One
is fixed at 60 μsec and one variable from 60 to about 40 μsec. Both signals are
converted from 101 to 30 Mc/s by means of a common local oscillator before
passing through the lines. This is necessary as the attenuation is excessive
at the higher frequency.

A difficulty with the use of a radio link is that the amplitude stability is not
particularly high. Since the amplitude of the interference pattern is a function
of the individual amplitudes of each signal, it is necessary to stabilize them

G

before mixing.   This is achieved by defining the pass band of each signal by two identical 30 Mc/s filters of approximately $\frac{1}{2}$ Mc/s bandwidth and then amplifying each signal to controlled equal levels of about 1 V.   This ensures that the noise powers generated in each aerial and preamplifier system are equalized, so that, if the aerial gains and the noise factors do not vary and there is no other source of noise in the system, the overall sensitivity of the equipment will be constant. After mixing, the combined signal is rectified immediately without further amplification.

A phase switching system is also employed, operating at a frequency of 2 kc/s.   While not essential for the operation of the equipment, it is convenient because it is difficult to obtain a very high degree of stability from the automatic gain controls.   A high switching frequency is desirable because the period of the interference pattern is very short with the longer spacings.

Some sample records obtained with the equipment on different sources and at different spacings are shown in Figure 2.

IV. Accuracy

There are many factors which can affect the accuracy of the relative amplitude measurements.   For convenience, the sources of error are considered under five headings :

(a) Time variations of sensitivity.

(b) Characteristics of the portable aerial sites.

(c) Confusion with other sources.

(d) Reading errors in the presence of random fluctuations.

(e) Systematic changes in sensitivity with distance.

The effect of the first four sources of error is to cause a more or less random dispersion of the measured amplitudes about their true values.   Examination of the results of Section V show that there are no obvious effects of this kind. It is estimated that their total effect results in a probable error in the relative amplitudes of the Cygnus source of less than 5 per cent. at the closer spacings and less than 10 per cent. at the 5 and 10 km spacings.   With the other sources the probable errors are thought to be less than 10 per cent.

The fifth source of error is particularly important because any systematic reduction in sensitivity as the aerial spacing is increased could result in an erroneous claim that a source had been resolved.   It is thought that errors of this type are negligible.

(a) Time Variations of Sensitivity

Here are included all those parameters of the system which can affect the overall sensitivity at a fixed aerial spacing and which tend to vary with the passage of time.   They include aerial gain, preamplifier sensitivity, signal-to-noise ratio on the radio link receivers, frequency pass bands, automatic gain control levels, and the sensitivity of the recorder and its amplifying system. The effect of such variables on the sensitivity was checked at intervals by returning the portable aerial to a fixed site and observing the amplitude of the pattern recorded on the Cygnus source.   In the absence of actual equipment faults, the maximum range in sensitivity was about 15 per cent., and usually

the deflexion was within 5 per cent. of the mean value taken as standard. A device was also developed which could be used for testing the overall sensitivity, excluding the aerial gain, without moving the portable equipment. It consists of a diode noise generator which is connected to the input of a preamplifier and pulsed at the switching frequency of 2 kc/s. The switch frequency is then adjusted until there is a difference of about $\frac{1}{4}$ c/s. The result is a beat pattern on the recorder at the difference frequency, the amplitude of which gives a measure of the sensitivity of the equipment.

(b) Characteristics of the Portable Aerial Sites

Variations in sensitivity at different sites can occur due to differences in the ground reflection effects and to differences in the propagation conditions over the radio link. Errors due to ground reflections were reduced to a minimum by selecting the various sites to be as flat and as free from obstructions as possible, and where necessary calculating corrections due to the ground reflected wave.

Providing the signal-to-noise ratio at the receiving end of the link is high it can have no effect on sensitivity. This was usually the case, but in one or two instances, particularly with the maximum spacing of 10 km where the signal-to-noise ratio was as low as 10 db, a correction had to be applied.

(c) Confusion with Other Sources

In addition to the source under observation, the aerials receive radiation from a large number of weaker sources. These will affect the relative amplitudes at each spacing in a random way. Errors of this type have been discussed by Smith (1952b), but his suggestion that a large number of observations will enable the error to be eliminated appears to be wrong, for there seems to be no way in which the component due to the wanted source may be separated, unless phases are measured.

(d) Reading Errors in the Presence of Random Fluctuations

A limit to the reading accuracy is set by the random noise fluctuations of the recorder, but they become important only when the relative amplitudes are low. Signal-to-fluctuation level at 1 km spacing ranges from about 5 for the Centaurus source to about 60 for the Cygnus source. A large number of recorder oscillations are averaged in every case to reduce this error to a minimum.

(e) Systematic Changes in Sensitivity with Distance

There are several factors which could lead to a reduction in sensitivity as the aerial spacing is increased. Because of their fundamental nature, errors of this type have received special attention and it is considered that the equipment is completely free from them, at least to a spacing of 5 km, and probably to many times this distance. Two possible sources of this kind of error in the equipment are phase instability, in which the variations in phase are more rapid than the recording system can follow, and frequency dispersion. Phase instability due to propagation changes and frequency dispersion both cause errors which tend to become worse as the length of the radio link is increased.

Possible phase instability was checked using 101 Mc/s transmissions from a small transmitter, which were received by both aerials and compared in phase

at the mixing point. This test was carried out with spacings of 1 and 5 km. It was found that in each case there were slow drifts in phase of up to 60° in an operating period of about $\frac{1}{4}$ hr, but that rapid drifts which could probably not be followed by the recorder in normal operation had an r.m.s. value of less than 3°. Because the results were similar on both spacings it is thought that the effects originated in the equipment rather than in the radio link. In any case they were so small as to have a negligible effect on the amplitude of the interference pattern.

The effect of dispersion was not expected to be important both because of theoretical considerations and because of practical experience with pulsed transmission systems. It was tested at the 5 km spacing by replacing the $\frac{1}{2}$ Mc/s filters defining the pass band of the equipment by filters with a bandwidth of about 140 kc/s. If dispersion were present and the total output power from the receiver were kept constant, the amplitude of the interference pattern due to a source would be increased with the narrower bandwidth. There was no significant difference in the amplitudes so that the effect of dispersion can be ignored.

A further possible source of error of this type could be due to the presence of ionospheric irregularities. As far as the effects of the irregularities are understood at present, however, it does not appear likely that they could produce any systematic effect. This was confirmed when observations were being made on the Cygnus source with spacings of both 5 and 10 km, when some records were steady in both phase and amplitude while others showed the same sorts of fluctuations as are commonly observed with closer spacings. Only the steady records were used to obtain the relative amplitudes of the source.

Finally, at the longer spacings the period of the interference pattern becomes so short (approximately 5 sec at 10 km) that allowance must be made for a reduction in the response of the recording system. This was overcome to a certain extent by using a shorter integrating time constant at these spacings. Values used were 2 sec out to 1·5 km and $\frac{1}{2}$ sec for the 5 and 10 km observations. The necessary small corrections were then calculated.

V. MEASUREMENTS AND ANALYSIS

Observations were made on the sources Taurus–A, Virgo–A, Centaurus–A, and Cygnus–A with nine different spacings in an east-west direction ranging from 60 m to 10 km and with three spacings at other azimuths. The east-west observations were sufficiently numerous to obtain distributions in that direction with an accuracy appropriate to the assumption of symmetry. Of the other observations one set has been made at an azimuth of 164° (east of north) and a spacing of approximately 1 km and the other two at an azimuth of 24° and spacings of 1 and 2 km. These are adequate to orient the major and minor axes of the sources approximately without giving very much information about their distributions in these directions. A few more observations would have been desirable, but the necessity of concluding the investigation within a definite time limit prevented any further observational work. The measurements on each source and their analysis will now be considered in turn.

(a) Taurus–A

The observations on this source are given in Table 1 ; also given is a measure-
ment made with a different interferometer (Mills 1952b) at a close spacing of
60 m which sets an upper limit to the value of the relative amplitude at this
spacing.

TABLE 1

OBSERVATIONS OF THE RELATIVE AMPLITUDE OF TAURUS–A WITH DIFFERENT AERIAL
SPACINGS AND AZIMUTHS

Azimuth of Interferometer Axis	Aerial Spacing (km)	n (wavelengths)	Relative Amplitude
90°	0·06*	20	<1·06
	0·29	100	1·00
	0·50	168	0·79
	0·79	265	0·77
	1·02	344	0·55
	1·25	420	0·47
	1·54	518	0·35
	5·35	1800	†
	10·01	3400	†
164°	0·94	192	0·72
24°	1·01	219	0·87
	2·10	455	0·53

* Observations made with different equipment.

† The amplitude of the interference pattern was less than the noise fluctuations
of the equipment.   The relative amplitude is less than about 0·1.

These values are plotted in Figure 3 (a) and a smooth curve drawn through
the east-west points to give the *amplitude-aerial spacing spectrum*.   Inspection
of the points obtained in the other directions shows that the source is not
circularly symmetric as its apparent extent is less when measured at the 24°
azimuth (since the amplitude falls off less rapidly).
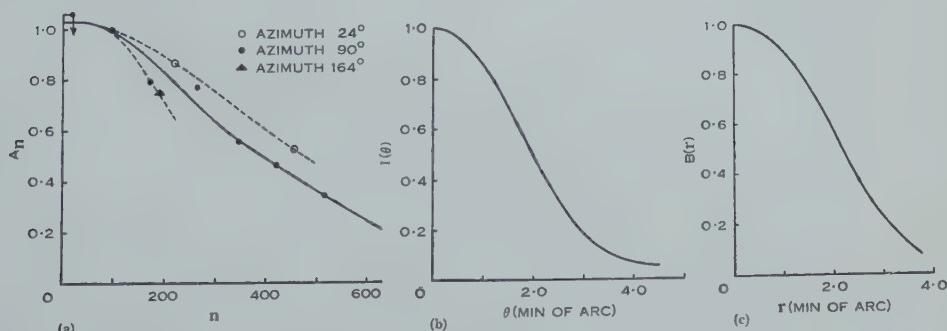
Fig. 3.—Taurus–A.

(a) Amplitude-spacing spectra.
(b) Integrated brightness distribution, E.-W.
(c) Maximum surface brightness distribution, E.-W.

In Figure 3 (*b*) is shown the integrated brightness distribution obtained
from equation (1) by putting φ_n equal to zero. The effective size defined before
(Mills 1952*c*) as the angle between the half-brightness points on this curve is
4·0′. In Figure 3 (*c*) is shown the equivalent radial brightness distribution
derived from equation (3). There is little difference between this and the
integrated brightness distribution.

The shape and orientation of the isophotes may be estimated from the
measurements made in the other directions. There is insufficient information
to construct the brightness distributions in these directions with any accuracy,
so the assumption is made that the shapes of the distributions are the same in
all directions. This places the further restriction on the isophotes that they
should have the same ellipticity ; it is consistent with the two observations
at an azimuth of 24°. The isophotes may now be constructed. From equations
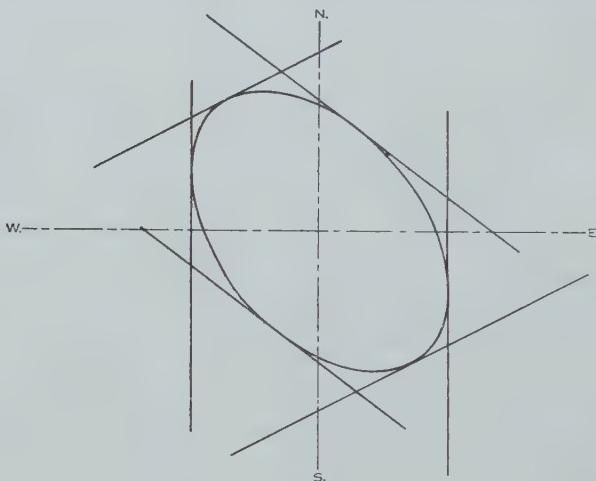
Fig. 4.—Construction of the half-brightness isophote of
Taurus–A.

(1) amd (3) and the description of method already given it will be seen that the
relative separation of the extremities of an isophote in any direction is inversely
proportional to the abscissae of the intersections of the appropriate amplitude
spectrum curve with any convenient ordinate. An ordinate of $A_n = 0·7$ has
been chosen, giving intercepts at $n = 265$ for the east-west distribution, $n = 210$
for an azimuth of 164°, and $n = 330$ for an azimuth of 24°.

Taking the half-brightness isophote as the most suitable one to construct
as it is least sensitive to variations in the amplitude spectrum, the separations
of the extremities of the isophote in these directions become 4·3′, 5·4′, and
3·5′ respectively. The position angles corresponding to the above azimuth
angles are then calculated from equation (4), and in Figure 4 the extensions are
drawn at right angles to the position angles and the elliptical isophote drawn
tangentially to these enclosing lines. The isophote has a size of $5\frac{1}{2}$ by $3\frac{1}{2}′$
with the major axis in position angle 140°.

The average central equivalent temperature of the source may now be estimated. On the basis of the dimensions above and the brightness distribution of Figure 3 (*c*), and by applying equation (5), this temperature is found to be $4 \times 10^6 \,^{\circ}\mathrm{K}$. Actual temperatures could exceed this by large amounts because a fine structure could exist and not be revealed by the present observations.
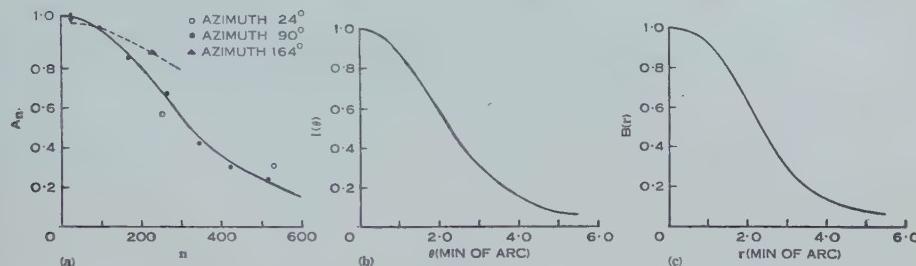
Fig. 5.—Virgo–A.

(*a*) Amplitude-spacing spectra.
(*b*) Integrated brightness distribution, E.-W.
(*c*) Maximum surface brightness distribution, E.-W.

(*b*) *Virgo–A*

The Virgo source has been treated similarly. The observational results are collected in Table 2 and the amplitude spectrum and the integrated and surface brightness distributions are plotted in Figures 5 (*a*)–(*c*). The east-west effective size is $4 \cdot 6'$.

TABLE 2

OBSERVATIONS OF THE RELATIVE AMPLITUDE OF VIRGO–A WITH DIFFERENT AERIAL SPACINGS AND AZIMUTHS

Azimuth of Interferometer Axis	Aerial Spacing (km)	*n* (wavelengths)	Relative Amplitude
90°	0·06*	20	<1·06
	0·29	100	1·00
	0·50	168	0·85
	0·79	265	0·67
	1·02	344	0·42
	1·25	420	0·30
	1·54	518	0·24
	5·35	1800	†
	10·01	3400	†
164°	0·94	226	0·88
24°	1·01	252	0·57
	2·10	525	0·31

* Observation made with different equipment.

† The amplitude of the interference pattern was less than the noise fluctuations of the equipment. The relative amplitude is less than about 0·1.

In obtaining the dimensions and orientation of the half-brightness contour there is a little difficulty, as it appears from the amplitude spectrum that the

distribution with the 24° azimuth is slightly different from the east-west distribution, showing a sharper concentration towards the centre and wider skirts.   However, the difference barely exceeds the accuracy of the observations so that the distributions have been assumed to be identical and the half-brightness ellipse drawn as before.   The size is then 5 by $2\frac{1}{2}'$ with the major axis in position angle 50°.

An estimate of the average central equivalent temperature gives 4×10^6 °K. Again the presence of fine detail could result in much higher temperatures.
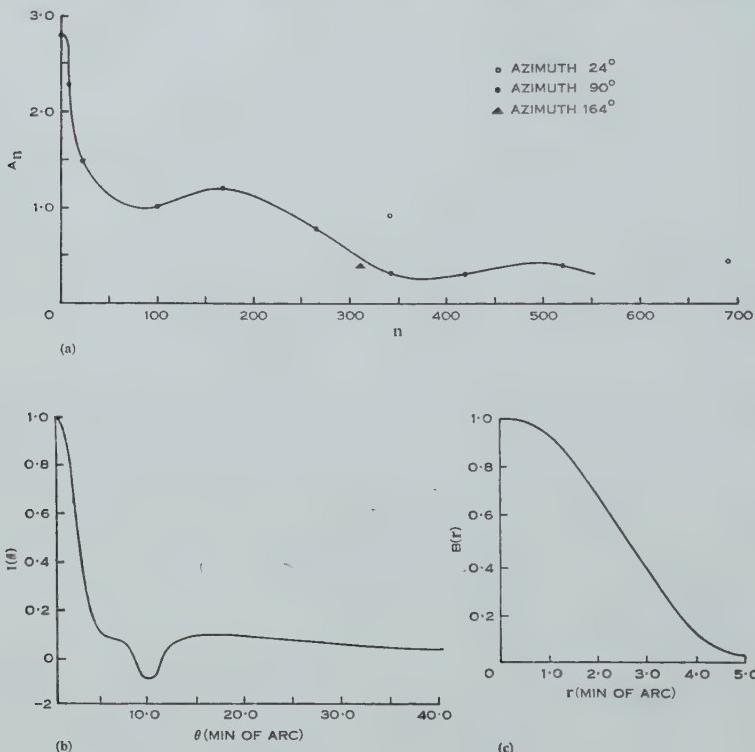
Fig. 6.—Centaurus–A.

(a) Amplitude-spacing spectra.
(b) Integrated brightness distribution, E.-W.
(c) Maximum surface brightness distribution, E.-W.

(c) Centaurus–A

This source presents special difficulties, as it has already been shown that it has a complicated distribution consisting essentially of an extended source of large angular size with a strong concentration near, but not coincident with its centre.*   It is now found in addition that the amplitude spectrum of the complex object has puzzling features.

In order to construct the east-west spectrum as accurately as possible, the present observations were supplemented by the results of two observations

---

* Reported by J. G. Bolton at U.R.S.I. General Assembly 1952.

with different equipments, (i) interferometer observations with a spacing of 60 m and (ii) an observation of the total intensity of the source obtained with the pencil-beam aerial described by Mills and Little (1953). The results of the observations are given in Table 3, and the amplitude spectra and the integrated and surface brightness distributions in Figures 6 (a)–(c).

An inspection of Figure 6 (b) reveals that over a small angle the derived integrated brightness is negative, which is a physical impossibility. The amplitude spectrum of Figure 6 (a) would therefore appear to be incorrect. A possibility that suggests itself is that the amplitudes at $n=420$ and $n=518$ could be negative, since no check is kept on their phases. A smooth curve could then be drawn through the points again, but this would lead to negative excursions

TABLE 3

OBSERVATIONS OF THE RELATIVE AMPLITUDE OF CENTAURUS–A WITH DIFFERENT
AERIAL SPACINGS AND AZIMUTHS

Azimuth of Interferometer Axis	Aerial Spacing (km)	n (wavelengths)	Relative Amplitude
90°	0*	0	2·80
	0·021	7	2·30
	0·060*	20	1·50
	0·29	100	1·00
	0·50	168	1·20
	0·79	265	0·80
	1·02	344	0·33
	1·25	420	0·33
	1·54	518	0·40
	5·35	1800	†
	10·01	3400	†
164°	0·94	313	0·40
24°	1·01	337	0·92
	2·10	700	0·40

* Observation made with different equipment.

† The amplitude of the interference pattern was less than the noise fluctuations of the equipment. The relative amplitude is less than about 0·2.

of brightness of much larger amplitude and is therefore unlikely to be correct. Some such anomaly is not surprising, however, as symmetry has been assumed and it may not exist. The result is salutary in demonstrating the errors which may arise owing to such an assumption. Another reason may be the omission of components due to spacings between 1·5 and 5 km, which could be appreciable. Because of the above anomaly the radial brightness distribution curve has been drawn only over the central concentration.

Two comparatively reliable results which can be obtained from the curves are that approximately 45 per cent. of the total energy of the source is in the central concentration and it has an effective size east-west of 5′. The effective size of the extended portion is difficult to estimate but it is of the order of $1\frac{1}{2}°$.

An attempt has also been made to construct the half-brightness isophote of the central concentration and estimate the central temperature as before. The same assumption of spectra of similar shapes in every direction is made which is consistent with the observed points. The size of the ellipse is then $6\frac{1}{2}$ by $3'$ with the major axis in position angle $130°$. While the accuracy is not expected to be very good, it seems quite clear that the extension of the source measured with an azimuth of $24°$ is much less than in the other directions.
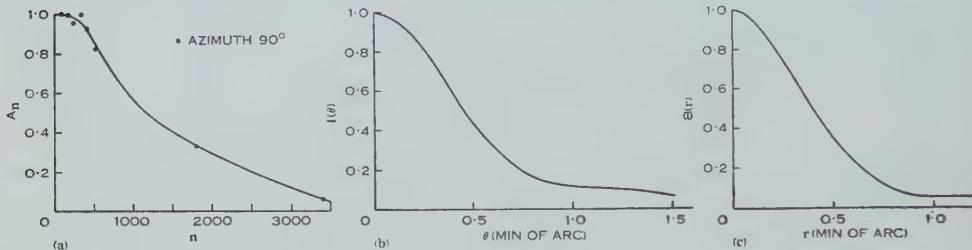
Fig. 7.—Cygnus–A.

(a) Amplitude-spacing spectrum.
(b) Integrated brightness distribution, E.-W.
(c) Maximum surface brightness distribution, E.-W.

Estimates of the equivalent temperatures give 4×10^6 °K for that of the central concentration and a value of the order of 3×10^4 °K for the brightest portions of the extended source (on the assumption of circular symmetry in the latter case).

(d) Cygnus–A

The observations on this source are tabulated in Table 4. Only the east-west observations are given as no significant changes in amplitude were detected in the other directions. This is not surprising as, with the large zenith angle of the source at Sydney, the effective aerial separations in these directions as given by equation (2) are very small.

TABLE 4

OBSERVATIONS OF THE RELATIVE AMPLITUDE OF CYGNUS–A WITH DIFFERENT
SPACINGS IN AN EAST-WEST DIRECTION

Aerial Spacing (km)	n (wavelengths)	Relative Amplitude
0·29	100	1·00
0·50	168	1·00
0·79	265	0·98
1·02	344	1·00
1·25	420	0·93
1·54	518	0·83
5·35	1800	0·33
10·01	3400	0·06

In Figure 7 (*a*) these points are plotted and a smooth curve is drawn through them. There are insufficient observations at large spacings to define this curve with certainty, and indeed it could well cross the zero axis several times. However, as drawn, it possesses the same features of a comparatively rapid initial fall and a long " tail " which is characteristic of the spectra of the other sources and in addition it is physically plausible.

The integrated brightness distribution $I(\theta)$ and the surface brightness distribution $B(r)$ are shown in Figures 7 (*b*) and 7 (*c*) respectively. From the latter curve can be calculated an average central equivalent temperature of the source on the assumption of circular symmetry. It is equal to 6×10^8 °K. Since the observations of Brown, Jennison, and das Gupta (1952) have shown that the Cygnus source is decidedly elongated with its longer axis not far from the east-west direction, the actual temperature is likely to be several times this value. Also it should be pointed out that the observations have been taken only to a distance of 10 km, so that the possibility of finer detail in the brightness distribution should not be overlooked.

VI. DISCUSSION

In Section V a certain amount of information about the properties of the sources was extracted from the observations. It now remains to consider these properties in relation to the nebulae with which the sources are identified and to try to draw some general conclusions. Various properties of the sources and related nebulae are tabulated in Table 5. A large part of the information concerning the nebulae is due to Baade and Minkowski (1953*a*, 1953*b*).

A further comparison between the Taurus, Virgo, and Centaurus sources and their respective nebulae is shown in Plate 2, where photographs of the nebulae are compared directly with " radio pictures " of the sources constructed from the data derived in Section V. While there is no particular reason why the radio isophotes should agree with optical features of the nebulae, it is clear that there are marked similarities, especially for two of the sources. The sources will now be discussed in turn.

First consider the Cygnus source and its associated nebula. Baade and Minkowski describe the nebula as having a bright central region of about 3 by 5″ surrounded by a much larger fainter part of elliptical outline, about 18 by 30″ with the major axis in position angle 150°. From Figure 7 it is seen that the radio source has an effective size in an east-west direction of about 45″. This is larger than the nebula but of the same order of size. Baade and Minkowski have shown that the nebula consists of two galaxies in collision, so presumably it is the excited interstellar gas which is responsible for the radio-frequency radiation, and it is not surprising that the gas should extend beyond the visible nucleus. Since it is unlikely that the temperature of the gas could rise as high as 6×10^8 °K or more as the result of a collision, it is probable that a non-thermal origin must be invoked for the radiation. This is also indicated by the spectrum of the source (Stanley and Slee 1950) which could not be ɟroduced by thermal emission from a hot, optically thin gas.

The Taurus source can be compared with its associated nebula, the Crab nebula, in rather more detail, as there is more information concerning the radio source and the nebula has been the subject of extensive investigations.   Here the optical and radio shapes are very similar.   It is well known that the nebula consists of an amorphous mass of gas at a temperature which has been estimated as 50,000 °K (Minkowski 1942), together with a filamentary network.   From the radio observations it is impossible to decide which of these features is responsible for the radiation, although it appears likely that it originates over

<div align="center">TABLE 5</div>

<div align="center">A COMPARISON BETWEEN THE RADIO SOURCES AND THEIR ASSOCIATED NEBULAE</div>

Radio source		Taurus–A	Virgo–A	Centaurus–A	Cygnus–A
Nebula		NGC 1952	NGC 4486	NGC 5128	—
Nebula type		Galactic, Supernova remnant	Extra-galactic, E_0	Extra-galactic, Peculiar	Extra-galactic, Two galaxies in collision
Magnitude	Radio m_R*	3·4	3·8	3·4	1·3
	Photographic m_p	9·0	9·9	7·2	17
Angular size	Radio source— effective size E.-W.	4·0′	4·6′	5·0′	45″
	Radio source ($\frac{1}{2}$ brightness)	5$\frac{1}{2}$ by 3$\frac{1}{2}$′	5 by 2$\frac{1}{2}$′	6$\frac{1}{2}$ by 3′ (Central concentration)	45″
	Nebula (approx.)	6 by 4′	5 by 5′	6 by 6′ (to $\frac{1}{2}$ brightness)	30 by 18″
Position angle	Radio source	140°	50°	130°	—
	Nebula	120°	—	—	150°
	Abnormal features of nebula	—	A " jet " at 290°	Obscuring band at 135°	—
Central equivalent brightness temperature of radio source		≈4×10⁶ °K	≈4×10⁶ °K	≈4×10⁶ °K	>6×10⁸ °K

* The magnitude system of Brown and Hazard (1952) is used here, i.e. $m_R = -53\cdot4 - 2\cdot5 \log_{10}S$, where S is the flux density in W m⁻² (c/s)⁻¹.

the whole space occupied by the nebula.   The central temperature of the source seems far too high to be explained in terms of thermal radiation from the mass of gas, although Stanley and Slee (1950) suggest that the spectrum is compatible with such an origin.   However, the latest results of Stanley, Slee, and Bolton (personal communication) contradict the previous results and suggest a spectrum more like that of the other sources.   It seems probable that a non-thermal process is operative in the gaseous mass of the nebula as also appears likely in the case of the Cygnus source.

The nebula NGC 4486 which is associated with the source Virgo–A is very peculiar.   It appears to be a typical E_0 galaxy which has the unique feature

RADIO BRIGHTNESS DISTRIBUTION OF COSMIC NOISE
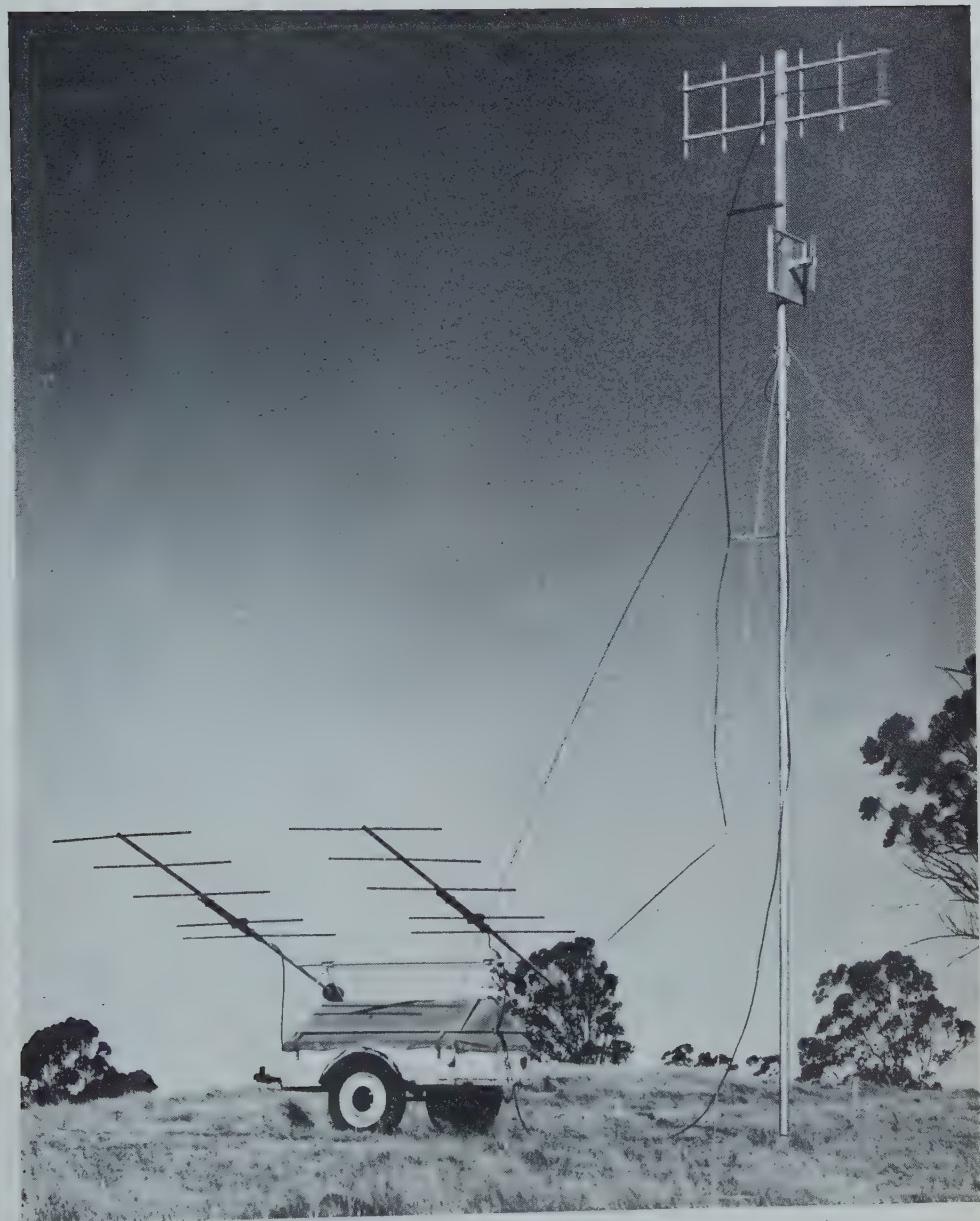
The portable aerial and associated equipment. Subsequently the receiving Yagi aerials were removed to a distance of about 10 m from the transmitting equipment to avoid troublesome interaction effects.

RADIO BRIGHTNESS DISTRIBUTION OF COSMIC NOISE
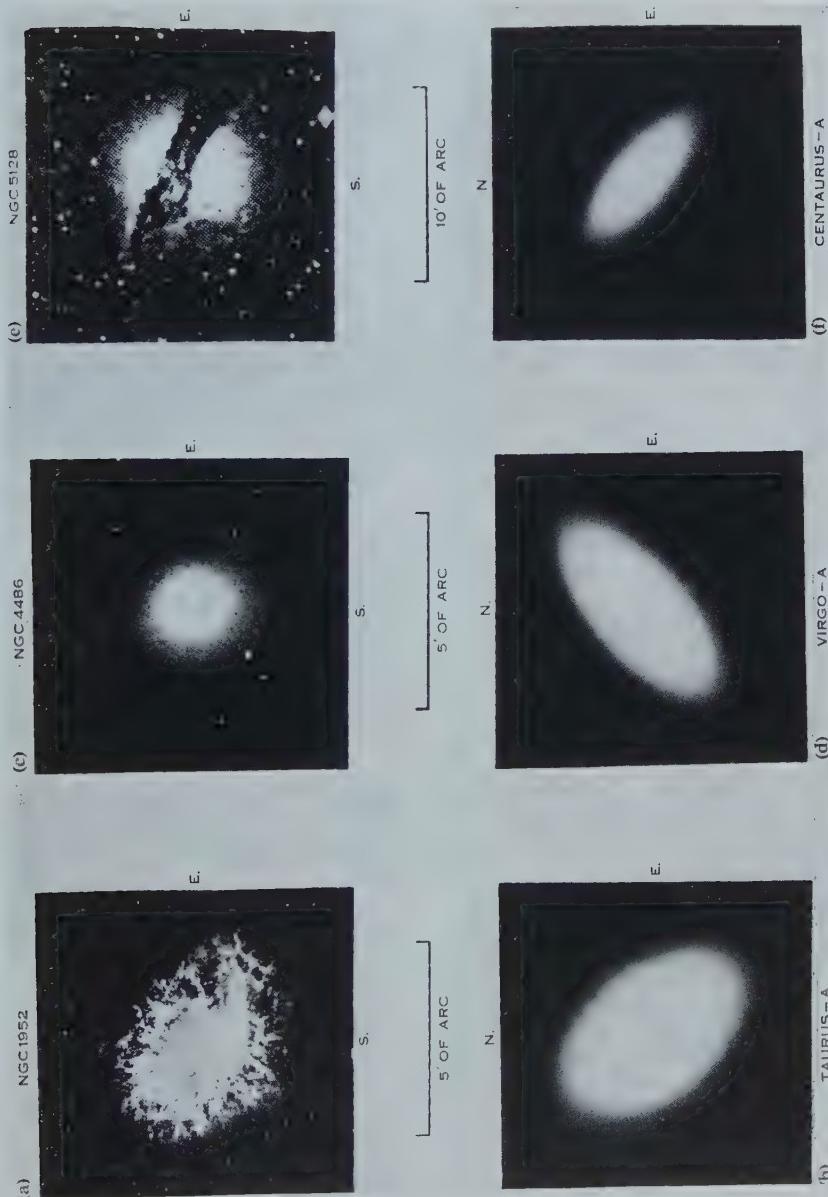
Photographs of the associated nebulae compared directly with "radio pictures" of the sources constructed from the observational data.

of a bright strip or " jet " of unknown composition crossing the nucleus at a position angle of about 290°. This jet is not visible in the photograph in Plate 2. Neither the nebula as a whole which is circular nor the jet which has a visible extent of only about 20″ of arc and is nearly at right angles to the major axis of the radio source agrees precisely with the shape of the source. However, the length of the major axis of the source is of the same order as the overall diameter of the nebula so that it appears likely that the radio emission is associated with the nebula as a whole rather than with the jet, although, with the present observations, the possibility that the radiation is produced by an irregular line source contained within the region of the assumed elliptical isophotes cannot be excluded.

The central concentration of the Centaurus source appears to be definitely associated with the central portion of the galaxy NGC 5128, and in particular with the obscuring band which crosses the galaxy. Since we might expect to find a concentration of gas in this region it would appear that this source also, along with Cygnus, Taurus, and possibly the Virgo sources, could have its origin in interstellar gas ; and again the temperature seems too high and the spectrum unlikely for a thermal origin. The extended source is rather more of a puzzle, but de Vaucouleurs (personal communication) has found that the nebula extends to a radius of at least 50′ so that an association between the extended source and the outer regions of the nebula is not unlikely.

In conclusion it might be pointed out that the surface brightness temperatures of three of the sources are similar, suggesting a possible origin in similar processes carried on at about the same level of intensity. The exceptionally high surface temperature of the Cygnus source is not surprising in view of the extreme effects which might be expected to occur in such a rare event as the head-on collision of two galaxies.

VII. Acknowledgments

VIII. References

BAADE, W., and MINKOWSKI, R. (1953a).—Identification of the radio sources in Cassiopeia, Cygnus-A, and Puppis-A. *Astrophys. J.* (in press).

BAADE, W., and MINKOWSKI, R. (1953b). On the identification of radio sources. *Astrophys. J.* (in press).

BOLTON, J. G., and STANLEY, G. J. (1948).—*Nature* **161** : 312.

BOLTON, J. G., STANLEY, G. J., and SLEE, O. B. (1949).—*Nature* **164** : 101.

BRACEWELL, R. N. (1952).—*Observatory* **72** (866) : 27.

BROWN, R. H., and HAZARD, C. (1952).—*Phil. Mag.* **43** : 137-52.

BROWN, R. H., JENNISON, R. C., and DAS GUPTA, M. K. (1952).—*Nature* **170** : 1061.

MACHIN, K. E. (1951).—*Nature* **167** : 889.

McCREADY, L. L., PAWSEY, J. L., and PAYNE-SCOTT, RUBY (1947).—*Proc. Roy. Soc.* A **190** : 357-75.

MILLS, B. Y. (1952a).—*Aust. J. Sci. Res.* A **5** : 266-87.

MILLS, B. Y. (1952b).—*Aust. J. Sci. Res.* A **5** : 456-63.
MILLS, B. Y. (1952c).—*Nature* **170** : 1063.
MILLS, B. Y., and LITTLE, A. G. (1953).—*Aust. J. Phys.* **6** : 272-8.
MILLS, B. Y., and THOMAS, A. B. (1951).—*Aust. J. Sci. Res.* A **4** : 158-71.
MINKOWSKI, R. (1942).—*Astrophys. J.* **96** : 199.
PIDDINGTON, J. H., and MINNETT, H. C. (1952).—*Aust. J. Sci. Res.* A **5** : 17-31.
SMITH, F. G. (1951).—*Nature* **168** : 555.
SMITH, F. G. (1952a).—*Nature* **170** : 1065.
SMITH, F. G. (1952b).—*Proc. Phys. Soc. Lond.* B **65** : 971-80.
STANIER, H. M. (1950).—*Nature* **165** : 354.
STANLEY, G. J., and SLEE, O. B. (1950).—*Aust. J. Sci. Res.* A **3** : 234-50.

APPENDIX I

Calculation of the " Integrated " Brightness Distribution

For a point source the normalized response of an interferometer in a direction near the collimation plane is given very closely by

$$R(n,\theta) = \cos(2\pi n\theta),$$

where n is the effective aerial spacing in wavelengths and θ is the angle between the source direction and collimation plane.

For a distributed source of " integrated " brightness $I(\theta)$, the response is obtained by applying R to I thus,

$$G(n,\xi) = \int I(\theta)\, R(n,\overline{\theta-\xi})\mathrm{d}\theta.$$

The integration is carried out over the primary aerial beam, but, since the angle of this beam is very much greater than the angular width of $I(\theta)$, the integration limits may be made infinite without sensible error.   We then have,

$$G(n,\xi) = \int_{-\infty}^{+\infty} I(\theta)\, \cos\,(2\pi n\overline{\theta-\xi})\mathrm{d}\theta.$$

Introducing complex notation,

$$A_n \exp\,(-\mathrm{i}\varphi_n)\, \exp\,(-\mathrm{i}2\pi n\xi) = \int_{-\infty}^{+\infty} I(\theta)\, \exp\,(\mathrm{i}2\pi n\overline{\theta-\xi})\, \mathrm{d}\theta,$$

where A_n is the relative amplitude and φ_n the relative phase of the interferometer output.

Eliminating $\exp\,(-\mathrm{i}2\pi n\xi)$ we have

$$A_n \exp\,(-\mathrm{i}\varphi_n) = \int_{-\infty}^{+\infty} I(\theta)\, \exp\,(\mathrm{i}2\pi n\theta)\mathrm{d}\theta,$$

whence

$$I(\theta) = \int_{-\infty}^{+\infty} A_n \exp\,(-\mathrm{i}\varphi_n)\, \exp\,(-\mathrm{i}2\pi n\theta)\mathrm{d}n$$

$$= 2\int_{0}^{\infty} A_n \cos\,(\varphi_n + 2\pi n\theta)\, \mathrm{d}n.$$

Only relative values of $I(\theta)$ are used in the analysis, so that the factor 2 may be omitted.

FLIGHT CHARACTERISTICS OF EXPANSIBLE BALLOONS

J. F. Darby,* V. D. Hopper,* Jean E. Laby,* and A. R. W. Wilson*

Summary

A general study is made of the factors governing the flight of rubber balloons for cosmic ray research, and a brief account is given of the methods that have been used to obtain long exposures at great heights. The results of incidental upper air wind observations in Victoria over a period of 18 months are included.

I. Introduction

Cosmic radiation investigations are being carried out at the Melbourne University by studying nuclear emulsion photographic plates which have been exposed to the radiation at high altitudes. It is desirable that the plates should spend a long time at altitudes greater than 80,000 ft and, moreover, have a high probability of being recovered.

Non-expanding balloons made of thin plastic sheet stabilize at a height fixed by their size and load and hence can give long exposures. However, such balloons are expensive and difficult to launch and all our work has been done with relatively inexpensive rubber balloons. Such a balloon inflated and sealed at the ground will rise at a nearly constant rate until it bursts. The present paper discusses the behaviour of these balloons together with methods of achieving long flights and improving recovery probabilities.

II. Rate of Rise

For values of Reynolds number met in balloon flights, the resistance to a rising balloon is determined by the transfer of momentum to the air (Gibson 1923); then for a spherical balloon of radius r, rising at a speed v, the free lift F in gravitational units is equal to the air resistance.

$$F = \pi k \rho v^2 r^2 / g, \quad \dots\dots\dots\dots\dots\dots (1)$$

where the drag coefficient k is a dimensionless quantity whose value depends on the Reynolds number R, πr^2 is the cross-sectional area of the balloon, and ρ is the density of the surrounding air.

If G is the gross lift, then the speed v is proportional to

$$F^{1/2}/k^{1/2}G^{1/3}. \quad \dots\dots\dots\dots\dots\dots (2)$$

It follows, as Mallock (1907) has shown, that a spherical balloon should rise at a speed proportional to the inverse sixth root of the density of the air surrounding it. Clarke and Korff (1941) have derived the equivalent result that the speed is proportional to the square root of the balloon radius.

* Physics Department, University of Melbourne.

Figure 1 shows some properties of the standard atmosphere of the International Committee for Air Navigation (Aeronautical Research Council 1952). Curve A shows the density relative to ground ρ/ρ_0, and curve B shows $(\rho/\rho_0)^{1/6}$, as functions of height. Integrating curve B (Fig. 1) gives the height reached by a balloon as a function of v_0t, where v_0 is the initial speed, and t the elapsed time, provided k remains constant.
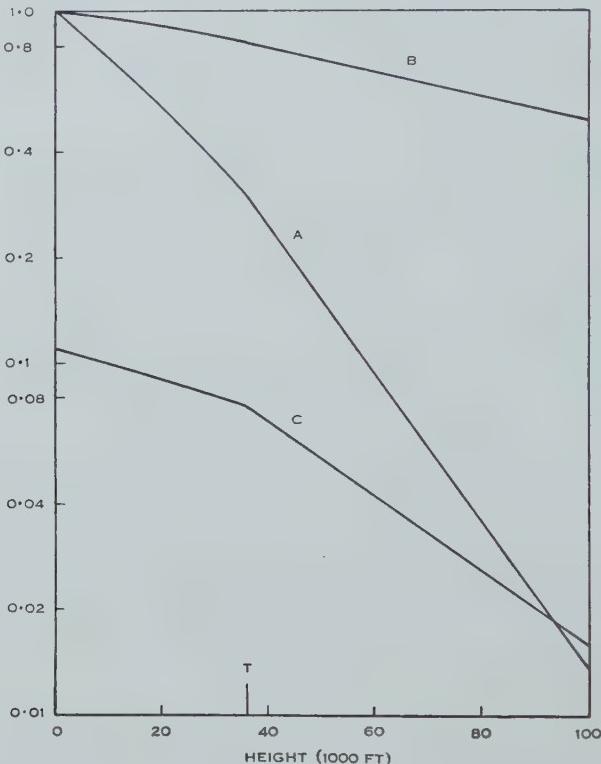
Fig. 1.—Curve A, density relative to ground ρ/ρ_0; curve B, $(\rho/\rho_0)^{1/6}$; curve C, $10^{-5}RF^{-\frac{1}{3}}$ gram$^{\frac{1}{3}}$ (R=Reynolds number, F=free lift) plotted against height for I.C.A.N. standard atmosphere. T is the tropopause.

Above a certain critical value of R, k is fairly constant and has a value near $0 \cdot 1$; at the critical value of R it increases suddenly and for lower values of R has a value k' near $0 \cdot 2$–$0 \cdot 3$. Therefore, if R decreases to the critical value during the flight, the resistance will increase suddenly and the speed will fall in the ratio of the square roots of the values of k.

Now Reynolds number

$$R = rv\rho/\eta$$
$$= (g\rho F/\pi k\eta^2)^{\frac{1}{3}}, \quad \dots\dots\dots\dots\dots \quad (3)$$

where η is the viscosity of the air and for balloons of common size R lies between 10^4 and 10^6 and decreases as the balloon rises. From relation (3), in the stratosphere where η is constant R is proportional to $\rho^{\frac{1}{3}}$.

Figure 2 shows the height v. time variation for a flight of a 2000 g balloon with a load of 1000 g and a free lift of 1000 g. The points are heights calculated from observations made with two theodolites and the curve is the theoretical height-time variation for an initial speed of 1100 ft/min. At 53,500 ft there is a sudden change of speed by a factor $0 \cdot 6$; this corresponds to a change of k in the ratio $0 \cdot 36$ which agrees with the values found by Ewald, Pöschl, and Prandtl (1923) in wind tunnel experiments. However, later work (Goldstein 1938) has shown fairly wide variations in k/k', values between $0 \cdot 4$ and $0 \cdot 2$ being found, and also in the critical value of Reynolds number which varies from 10^6 to 10^5. In Figure 1 curve C, $R/F^{\frac{1}{2}} = (g\rho/\pi k\eta^2)^{\frac{1}{2}}$ is plotted for $k = 0 \cdot 1$; at 53,500 ft with
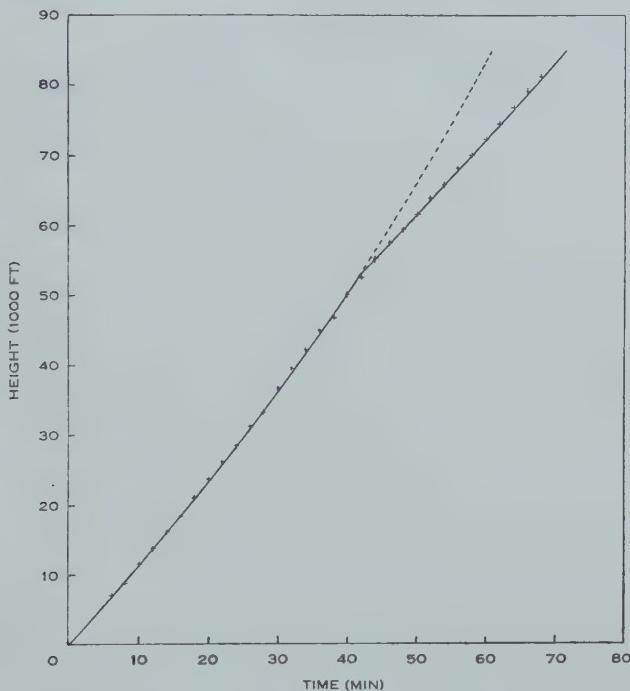
Fig. 2.—Flight of a 2000 g balloon. Solid line shows theoretical height with k changing by $1/0 \cdot 6$ at 53,500 ft; crosses are observed heights. Dotted line shows expected height if k does not change.

free lift 1000 g the calculated critical value of R is $1 \cdot 5 \times 10^5$. Two other flights that have been analysed in detail have given speed changes by factors of $0 \cdot 7$ and $0 \cdot 9$ at Reynolds numbers of $1 \cdot 9 \times 10^5$ and $1 \cdot 6 \times 10^5$ respectively, which are in the range mentioned above. The result of this investigation has therefore shown that a variation of rate of rise does occur in flights with these large balloons which is in general agreement with theory. For most purposes the standard method of assuming a constant rate of rise is sufficiently accurate.

III. Excess Pressure in Spherical Balloons

Owing to the tension in the rubber, the pressure inside a rubber balloon exceeds the external pressure when the diameter is greater than the flaccid

diameter. This excess pressure must not be neglected when considering the small free lifts required for nearly stabilized balloons at great heights. If it is assumed that Poisson's ratio for the balloon rubber remains at $0 \cdot 5$ and Young's modulus stays constant throughout the expansion of the balloon, then the excess pressure Δp at radius r is given by

$$\Delta p = 0, \qquad\qquad\qquad \text{for } n \leqslant 1,$$
$$\Delta p = (CM/r_f^3)(n-1)/n^3, \quad \text{for } n > 1, \quad \dots\dots\dots\dots \text{(4)}$$

where r_f = flaccid balloon radius,

$n = r/r_f$,

M = mass of balloon rubber,

C = constant depending on the elastic moduli of the rubber and its density.
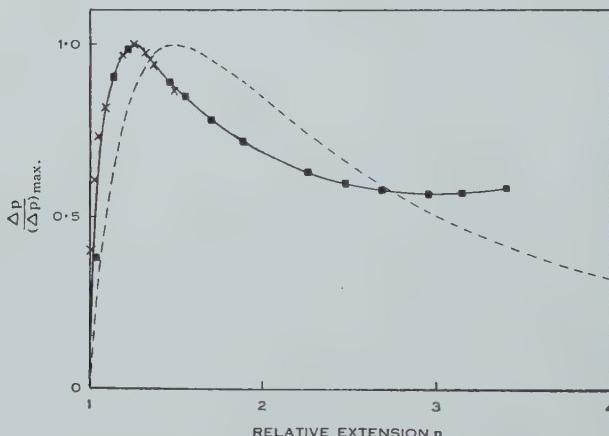
Fig. 3.—Ratio of excess pressure to its maximum value as a function of relative extension n. Squares are experimental values for a 700 g balloon, crosses for a 2000 g balloon. Dotted curve theoretical values from equation 4. $\Delta p_{\max.} = 18 \cdot 3$ mm H_2O for 700 g, $7 \cdot 9$ mm H_2O for 2000 g.

Figure 3 compares the variation of excess pressure obtained from the above expression with experimental values for two sizes of " Beritex " balloons made by the Guide Bridge Rubber Company. The excess pressures were measured with a micrometer water manometer. As might be expected from the assumptions used in its derivation, relation (4) does not usefully represent the variation of excess pressure with relative inflation, particularly in that it fails to predict a further rise after the initial maximum.

Väisälä (1937) has previously measured the excess pressure variation inside balloons of less than 70 cm flaccid diameter. He found that the excess pressure Δp is given by

$$\Delta p = 2d_0 P(n)/r_f, \quad \dots\dots\dots\dots\dots\dots \text{(5a)}$$

where d_0 is the thickness of unstretched balloon rubber. $P(n)$ is characteristic of the rubber and is a complicated function empirically determined to be given by

$$P(n)=(\tau/n^3)\ \exp\ [a(n-1)-b/(n-1)],\ \ \ldots\ldots\ldots\ (5b)$$

τ, a, and b and hence the shape of the excess pressure curve are constant for all balloons made of the same rubber. This agrees with our experimental results since curves of $\Delta p/(\Delta p)_{max.}$ for the 700 and 2000 g balloons coincide (Fig. 3).

Both relations (4) and (5) give the ratio of the excess pressures at the same relative extension of two balloons made from the same rubber as

$$\frac{\Delta p_1}{\Delta p_2}=\frac{M_1}{M_2}\left(\frac{r_{f2}}{r_{f1}}\right)^3.\ \ \ldots\ldots\ldots\ldots\ldots\ldots\ (6)$$

For the balloons used in the present experiment this gives

$$\frac{\Delta p_{2000\ g}}{\Delta p_{700\ g}}=0\cdot42.$$

The experimental value was $0\cdot44$.

IV. Height Stabilization of Rubber Balloons

In order to keep a balloon at great heights over long periods, it is necessary that its free lift be made very close to zero at the required height. Because of the excess pressure inside a rubber balloon its free lift is not quite independent of height but diminishes slowly. This effect can be used to fix the ceiling height of a balloon by giving it a small initial free lift which can be calculated by reference to Figure 3 and is reduced to zero at the required ceiling height. However, the initial free lift needed to do this is small ($\sim$40 g for a 700 g balloon$+$400 g load to stabilize at 80,000 ft) so the rate of rise would be very slow and the time necessary to reach the proposed ceiling height might exceed the balloon's life, or the wind might take it out of sight before the ceiling height is reached. Also it is difficult to set the free lift accurately to the required value. A fairly rapid rise to the ceiling is wanted and the free lift must then be removed. We have tried to do this in the following ways :

(i) A commonly used method is to have strings of balloons which burst in succession until the free lift becomes negative ; then the remaining balloons carry the load slowly back to ground. With this method there is always something to sight on and to mark the load after falling, but one cannot know how much rubber from the burst balloons will remain as load. This is serious when few balloons are used.

(ii) The disadvantage of method (i) can be overcome by placing one balloon inside another. Then when the outer one bursts, the inner one is left unencumbered. This restricts the size of the load to what can be forced through the neck of a balloon ; also it was found that there was a danger of the inner balloon bursting at the same time as the outer one.

(iii) Another way to separate two balloons completely is to use a release mechanism, operated by a clock, and then the load can be bulky. In practice,

the clock was set so that when the pair of balloons had reached about 70,000 ft the upper one which had provided most of the lift was disconnected and the lower balloon, which carried the plates, rose more slowly until it burst. In methods (ii) and (iii) the free lift of one of the balloons must be set so as to have a low value at the ceiling, and this was found to be difficult to arrange in practice.

(iv) A simple way to remove the excess free lift at a preset height is to release the corresponding amount of gas by means of a valve operated by a string attached across a diameter (Hopper and Wilson 1953). A wide tube is inserted in the neck of the balloon ; we have found a diameter of $1\frac{1}{4}$ in. satis-
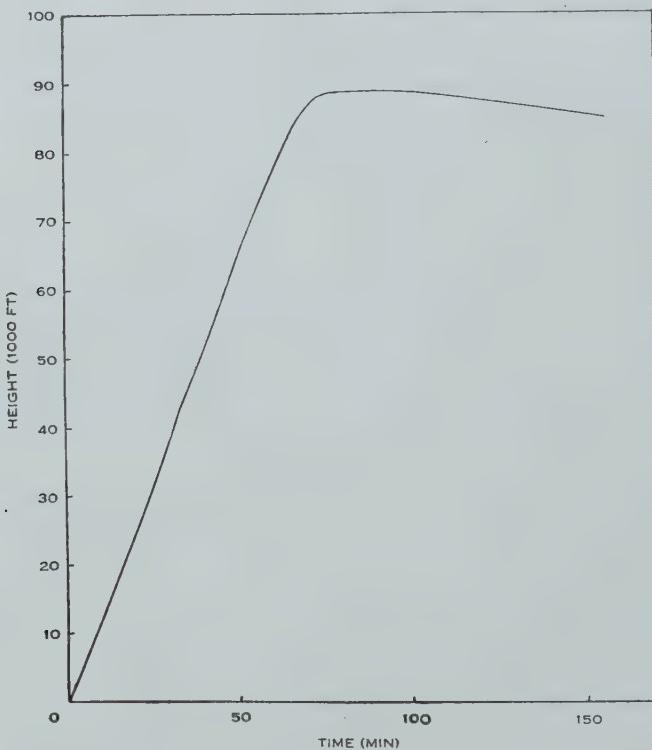
Fig. 4.—Flight of a 2000 g balloon with string-operated valve.

factory. A valve rests on top of the tube and a string from the valve is attached to the opposite end of the diameter of the balloon. When the balloon reaches a height at which the string becomes taut the valve opens and gas is released until the gross lift equals the load. Figure 4 shows a height-time curve for a flight of this kind.

(v) Any levelling-off process prevents the balloon from reaching its maximum (bursting) height. If maximum height as well as a long flight is wanted, the balloon must retain a small positive free lift. This can be achieved by first using the levelling-off process described in (iv) with a length of string such that the balloon levels off at a diameter below its bursting diameter.

When the balloon has levelled off, a clock closes the valve and then drops a small weight. This will cause the balloon to rise again until it reaches its bursting diameter or until the tension in the rubber causes it to level off at a new height.

This new height can be found in the following way. Assuming that the balloon has levelled off at a radius r_1, pressure p_1, and gross lift G_1 equal to the total load, then

$$G_1 = \frac{4}{3}\pi r_1{}^3(\rho_1^A - \rho_1^H),$$

where ρ_1^A is the density of the air at pressure p_1,
ρ_1^H is the density of hydrogen at pressure p_1.
The pressure of hydrogen is in fact $(p_1 + \Delta p_1)$ but $\Delta p_1 \ll p_1$ at all heights considered. If the valve is now closed and a small weight dropped, the load is changed to G_2, and the balloon will stabilize at a new pressure level p_2, with radius r_2 such that

$$G_2 = \frac{4}{3}\pi r_2{}^3(\rho_2^A - \rho_2^H),$$

where ρ_2^A and ρ_2^H are densities at the new pressure.

Then

$$\frac{G_1}{G_2} = \frac{r_1{}^3}{r_2{}^3} \cdot \frac{\rho_1^A - \rho_1^H}{\rho_2^A - \rho_2^H} \approx \frac{r_1{}^3 \rho_1^A}{r_2{}^3 \rho_2^A} = \frac{r_1{}^3 T_2 p_1}{r_2{}^3 T_1 p_2}, \quad \dots\dots\dots \text{(7)}$$

where T_1, T_2 are temperatures at the two heights.

Now considering the hydrogen

$$(p_1 + \Delta p_1)r_1{}^3/T_1 = (p_2 + \Delta p_2)r_2{}^3/T_2.$$

Substitute in (7)

$$(p_1 + \Delta p_1)T_2/(p_2 + \Delta p_2)T_1 = G_2 p_1 T_2/G_1 p_2 T_1,$$

or

$$\Delta p_2/p_2 = G_1(1 + \Delta p_1/p_1)/G_2 - 1.$$

From Figure 3, Δp is approximately constant for $n > 2$. Hence

$$p_2 = \Delta p_1/[G_1(1 + \Delta p_1/p_1)/G_2 - 1].$$

Loss of gas through expanded pin-holes which sometimes occurs in balloon rubber has so far prevented us from stabilizing a balloon at a new height after dropping a weight.

V. Tracking

We have followed balloons usually by the double theodolite method and occasionally, through the courtesy of the Commonwealth Meteorological Branch, by radar. On good days balloons can still be seen 90 miles or so away, though

they would usually be lost in haze at much shorter distances. To follow a complete flight it is desirable that the balloon should not recede more than about 50 miles. Fortunately it has been found that the wind structure in Victoria in clear weather in spring and summer often allows this. Figure 5 shows average winds that we have observed in flights over a period of 18 months. We have plotted westerly and southerly components and standard deviation following Brooks, Durst, and Carruthers (1946). The easterly component above 70,000 ft during spring and summer has also been observed by the Commonwealth
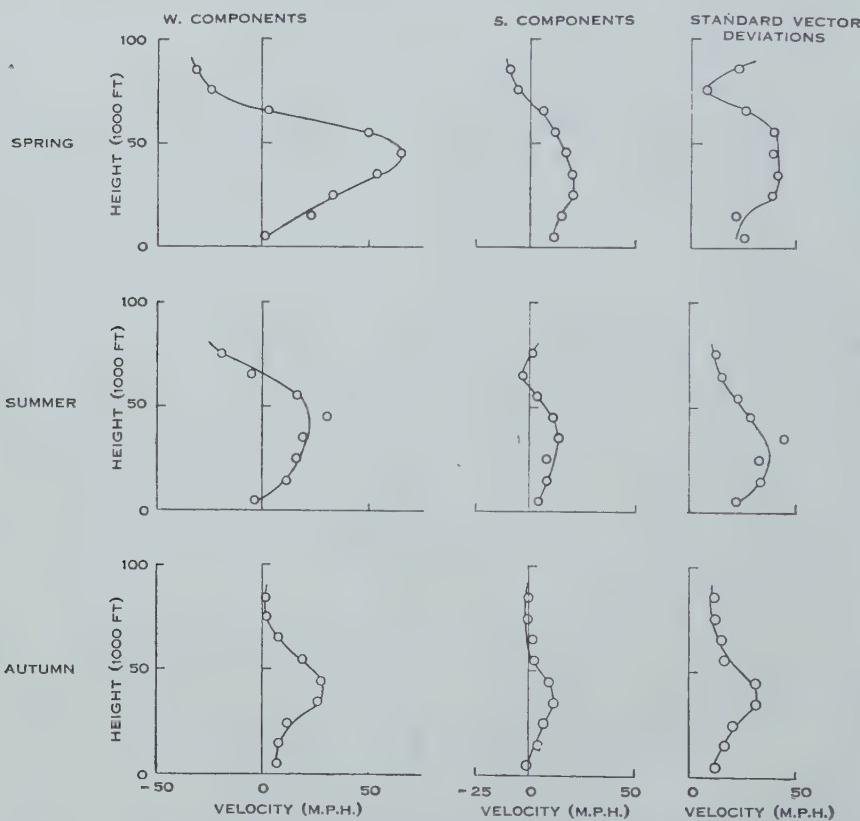
Fig. 5.—Seasonal mean winds and standard vector deviations as functions of height.

Meteorological Branch over South Australia (Trefry, personal communication 1952) and a similar pattern by the New Zealand Meteorological Service over Nandi and Auckland (Porter 1952).

Often from 20,000 to 60,000 ft the wind is between SW. and W. and above 70,000 ft between E. and NE., so the rising balloon travels east and returns. A stabilized balloon may remain so long in the easterly wind at its maximum height that it travels further than it did in the stronger westerly below. Figure 6 shows the path followed by the balloon in the flight shown in Figure 4. The dotted part is deduced from the wind structure disclosed by the observed part

of the flight. This return also helps recovery, as eastern Victoria is mountainous and thinly populated. Even so it is best to attach a marker that is easily seen and looks strange enough to attract the passer-by. A sphere made of a light frame covered with aluminium foil was first introduced by Mr. G. R. Trefry of Commonwealth Meteorological Branch and is quite suitable. In sunshine it can be distinguished several miles away and the recovery rate in Victoria since using such markers has been 29 out of 33 balloons, whereas previously it was 25 per cent.

Fig. 6.—Ground projection of track of a balloon, stabilized by a valve, released at A, observed to B, and found at C. Solid curve is observed path, dotted curve path deduced from observed winds. Graduations on curve are times from release in minutes.

VI. Acknowledgments

The work described in this paper summarizes a programme of work carried out by a team engaged in cosmic ray research at the University of Melbourne. Other members of the team who have contributed to preparing balloon flights and taking observations are Dr. R. Parsons, Dr. S. Biswas, Miss F. Brisbout, and recently Mr. I. Macaulay and Mr. Y. K. Lim. Special thanks must be given to Mr. H. A. Waters for technical assistance.

The continual support and interest of Professor L. H. Martin is gratefully acknowledged. The authors wish also to thank the Commonwealth Meteorological Branch and the Department of Supply for their generous help.

VII. References

Aeronautical Research Council (1952).—" Compressible Air Flow Tables." p. 140. (Oxford Univ. Press.)

Brooks, C. E. P., Durst, C. S., and Carruthers, N. (1946).—*Quart. J. R. Met. Soc.* **72** : 55.

Clarke, E. T., and Korff, S. A. (1941).—*J. Franklin Inst.* **232** : 339.

Ewald, P. P., Pöschl, T. L., and Prandtl, L. (1923).—*Ergebn. Aerodyn. VersAnst. Göttingen* **2** : 28.

Gibson, A. H. (1923).—" The Mechanical Properties of Fluids." Ch. 5. p. 198. (Blackie and Son Ltd. : Glasgow.)

Goldstein, S. (1938).—" Modern Developments in Fluid Dynamics." p. 495. (Oxford Univ. Press.)

Hopper, V. D., and Wilson, A. R. W. (1953).—*J. Sci. Instrum.* **30** : 211.

Mallock, A. (1907).—*Proc. Roy. Soc.* A **80** : 530.

Porter, Elizabeth M. (1952).—N.Z. Met. Serv. Tech. Note 92 (mimeo.).

Väisälä, V. (1937).—*Ann. Acad. Sci. Fenn.* A **48**, No. 8.

BOREHOLE TEMPERATURE MEASURING EQUIPMENT AND THE GEOTHERMAL FLUX IN TASMANIA

By G. Newstead* and A. Beck†

Summary

An instrument, using a thermistor and an A.C. bridge circuit, is described by which temperature measurements in boreholes can be made to $0 \cdot 01$ °C. Results of temperature measurements in Tasmania are given and a value of 2 cal cm^{-2} sec^{-1} is obtained for the heat flux (uncorrected for recent glaciation).

I. Introduction

The object of this paper is, firstly, to describe equipment for measuring the temperature in boreholes, which was developed by one of us (G.N.) as part of a programme of research on remote-reading bridges, and, secondly, to give results for geothermal gradients and heat flux obtained by its use.

The measuring element of the instrument is a thermistor connected by a polythene cable to an A.C. bridge situated at the top of the hole. Thermistors offer obvious advantages for temperature measurement since, because of their high temperature coefficients of resistance (of the order of 4 per cent.), the accuracy with which it is necessary to measure the resistance for a given accuracy of temperature measurement is much less than with other types of resistance thermometry. They have been used previously (Deeter 1948 ; Misener, Thompson, and Uffen 1951 ; Standard Telephones and Cables Ltd. 1952) in D.C. bridge circuits ; the novel feature of the present equipment is the use of an A.C. bridge. This was chosen mainly because of the ease of obtaining amplification. From the figures given in the paper it will be seen that to detect a temperature change of $0 \cdot 01$ °C requires the measurement of a current of 10^{-9} A which is about two orders of magnitude less than can be measured with available portable galvanometers. Electronic equipment has the advantages of being more portable, more robust, and more easily obtained than a special D.C. galvanometer. The use of D.C. amplifiers is not desirable either under field conditions or for the small voltages in question. Finally, the use of alternating current has other advantages, such as avoiding errors due to thermal and contact e.m.f.'s. It is found that, under typical field conditions, the absolute accuracy of the equipment is of the order of $0 \cdot 01$ °C or better.

The whole of the measuring apparatus is mounted in a suitcase and can be carried in the field. It may be run from either a car battery or dry batteries.

* University of Tasmania.

† Australian National University.

While it was designed primarily for temperature measurements in boreholes, it is hoped to use it also for geothermal prospecting and for temperature measurements in mines and tunnels.

No measurements of the geothermal flux have yet been made in Australia, and it was felt that Tasmania would be a particularly interesting field for them since it has been recently glaciated and has some boreholes through fairly homo-
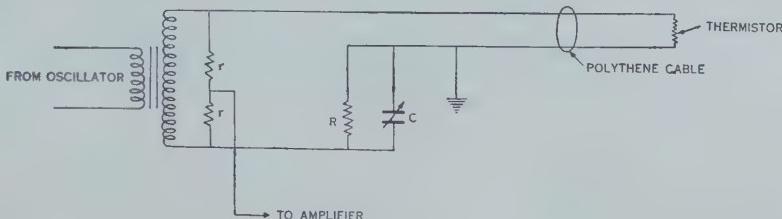
Fig. 1.—Schematic diagram of bridge.

geneous strata. It was hoped that, with the above accuracy in the temperature measurements, it might be possible to make a contribution towards the dating of the glaciation. In the event, however, many of the holes proved to be incomplete, unsuitable, or blocked, so that for the present it is only possible to give a few values for the heat flux.
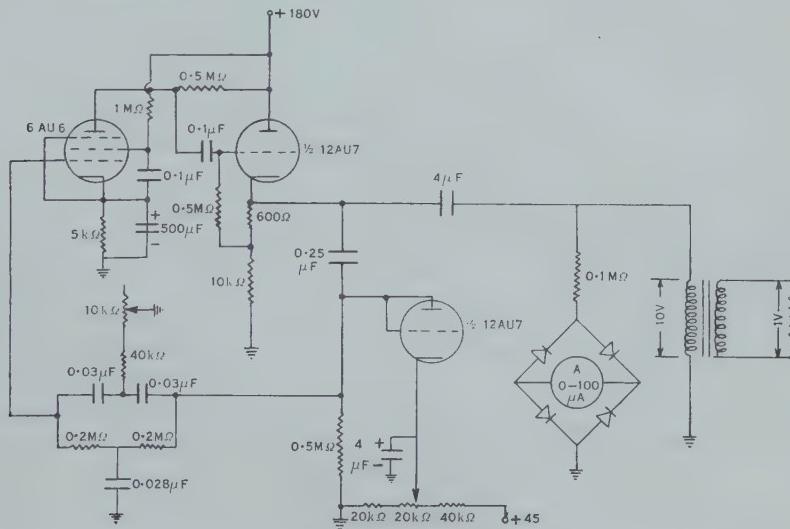
Fig. 2.—Oscillator circuit.

II. Temperature Measuring Equipment

A schematic circuit diagram is shown in Figure 1. The resistance R is a five dial Muirhead box, type A25U, the arms r are two 100 Ω " Minalpha " resistances, and the variable condenser C is built up from good quality mica dielectric condensers with air dielectric trimmers. The thermistor is an S.T.C. type A. Several values of resistance were used, but type A1522/100, which has a nominal resistance of 10^5 Ω at 20 °C, was found best to cover the range of

temperature in the boreholes. The cable was a Telcon type PT11 with about 1500 ft always in circuit.

The requirements for the oscillator are constancy of output and reasonable frequency stability and waveform. The frequency of 40 c/s was chosen in order to make the capacity currents in the cable small compared with the currents in the thermistor. A bridged T feedback oscillator with diode output stabilization as described by Lynch and Robertson (1946) and shown in Figure 2, was used. The oscillator develops 0·8 V across the bridge and a built-in voltmeter enables this voltage to be monitored. Controls are provided to set the output voltage and to vary the frequency by a small amount so as to centre it in the amplifier pass band.
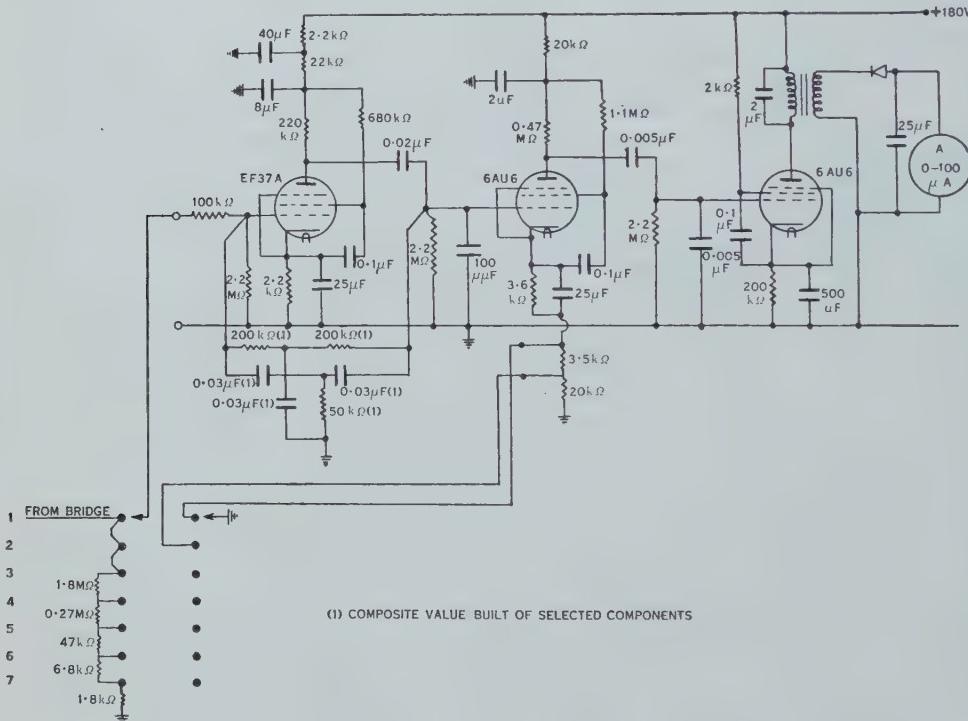
Fig. 3.—Amplifier circuit.

The circuit diagram of the amplifier is shown in Figure 3. The first stage incorporates a bridged T feedback circuit (Valley and Wallman 1948) to make it very selective : the resonant frequency is 40 c/s and the pass band is only a few cycles centred about this. An EF37A pentode is used to cut down microphonic effects. The design of the last two stages is quite conventional. The first two steps of the attenuator reduce the gain of the second stage by introducing cathode degeneration, and the next steps attenuate the input ; the reduction of gain is approximately seven times per step. Bridge balance is indicated as a minimum on the 0–100 μA meter which receives the rectified output of the third stage. On full sensitivity, 2·5 μV input gives full-scale deflexion ; this sensitivity is

more than required and allows for deterioration of the amplifiers as the batteries run down and the valves age. The input impedence at 40 c/s is 10^5 Ω and the amplifier noise produces a standing current of 5 μA. The power consumption of the whole equipment is 0·6 A at 12 V and 15 mA at 180 V.

The 1500 ft of cable, which covers the depths of the boreholes investigated, is wound on a reel; because of the danger of overstressing the cable if the measuring head sticks, the head is supported by a thin wire rope wound on a separate reel and run over a measuring wheel connected with a counter which reads depth directly. The cable and wire rope were clipped together every 100 ft. As an additional check on measured depths the wire rope was marked at 50-ft intervals.

Details of the temperature measuring head are shown in Figure 4. The essential problem in the design of the head is the provision of a water-tight seal on the cable which will not strain its dielectric. This has been effected by modifying a method of Munroe and Penny (1952). A polythene plug P is
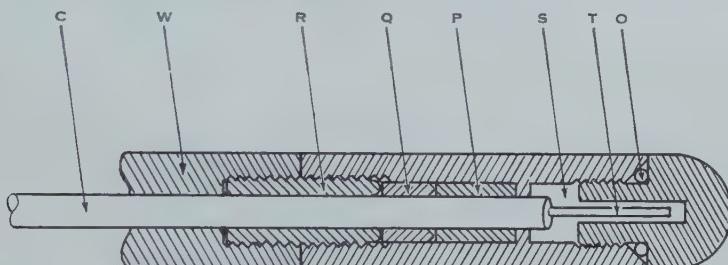
Fig. 4.—Temperature measuring head.

compressed by a ring Q which is forced in by a screw R until the polythene flows, making a seal with the cable C; the necessary setting of the screw R was found by preliminary experiments. These seals are easily made and have been tested in the laboratory to over 1000 lb/in.². This seal provides a watertight compartment S in which the thermistor T is situated. To facilitate assembly, which is frequently done in the field, the end of the compartment is removable and is sealed by a rubber O-ring at O; a cover, not shown in the figure, protects the thermistor from damage when removing the end. A brass weight W brings the total weight of the head up to about 12 lb. The diameter of the head is 1 in.

Enclosing the thermistor in this way implies that it requires several minutes to attain the temperature of its surroundings. This time lag has not proved to be a disadvantage, since in some earlier experiments with a thermistor in contact with the water it was found that fluctuations of temperature (presumably due to the stirring of water when lowering the head) occurred and took some minutes to die down.

III. Theory of the Method

The balance condition of the bridge requires that the capacity C and resistance R be adjusted to be equal to the capacity and resistance in the cable

arm. The capacity balance is completed first ; it is found that this does not alter appreciably as the cable is unreeled, indicating that the effect of the strains which occur in the cable in practice is negligible.

The thermistor with the cable in circuit was calibrated in the laboratory against a standardized mercury-in-glass thermometer (equipment for platinum thermometry not being available in Tasmania). The method is thus one of substitution and many sources of error are eliminated ; its performance, however, is limited by the stability of the thermistor. The makers' data indicate that their products can be relied upon to $0 \cdot 01$ °C and Doucet (1951) gives the limits of a similar type as $\pm 0 \cdot 002$ °C, but there is very little published information available on this point. In view of this we have adopted the practice of quickly checking the calibration at two or three points at each return to the laboratory. In all cases when a complete recalibration has been made, the original values have been reproduced to the required order of accuracy.

The complete calibration was carried out in a bath which was thermostatically controlled to $0 \cdot 01$ °C. Measurements were made at temperature intervals of about 1 °C, interpolation between these points being made by assuming the resistance R_T of the thermistor at temperature T to be of the form

$$R_T = a(T)e^{b(T)/T},$$

where $a(T)$ and $b(T)$ are slowly varying functions of T.

To determine the sensitivity and accuracy of the equipment, the theory of the resistance balance must be examined more closely. The resistance R_T of the thermistor is of the order of 10^5 Ω at 20 °C and its temperature coefficient α_T is of the order of $0 \cdot 04$. Thus a change in temperature of $0 \cdot 01$ °C causes a change in resistance of 40 Ω and applies a voltage of 50 μV to the amplifiers, which are thus seen to have adequate sensitivity.

The series resistance of the cable is approximately 14 Ω per 1000 ft of lead and return : the constant part of this is eliminated by the method of calibration, and the change in series resistance caused by likely variations in temperature will be less than 2 Ω with 1500 ft of cable in circuit. Thus the series resistance of the cable and the effect of temperature on it may be disregarded.

The effect of the shunt resistance of the cable and its variation with temperature and strain is more difficult to estimate. Taking the capacity of 1500 ft of cable as $0 \cdot 03$ μF, 40 c/s for the frequency, and 3×10^{-4} for the power factor of polythene gives 450 MΩ for the shunt resistance R_s of the cable.

Since

$$\frac{1}{R} = \frac{1}{R_s} + \frac{1}{R_T},$$

the change δR in R due to a change δR_s in R_s is

$$\delta R = \frac{R^2}{R_s^2} \delta R_s,$$

and the change δR due to a 30 per cent. change of shunt resistance for the case in which $R \simeq R_T \simeq 10^5$ Ω and $R_s = 4 \cdot 5 \times 10^8$ Ω, is about 7 Ω, which again is negligible compared with the 40 Ω change in R due to a change of $0 \cdot 01$ °C.

Another possible source of error is the heating of the thermistor by the voltage applied to it. The dissipation constant of the thermistor is $0 \cdot 28$ mW/°C and at balance the voltage applied to it is $0 \cdot 4$ V. Taking $10^5 \, \Omega$ as the resistance of the thermistor, we get a temperature rise of $0 \cdot 005$ °C due to the applied voltage. That this effect is not causing error can be demonstrated by varying the voltage applied to the bridge and noting that the balance resistance setting is unchanged ; if heating of the thermistor were causing false readings the balance resistance would be a rapidly changing function of the applied voltage. Nevertheless, the voltage in the thermistor is maintained at a constant value during the calibration and the measurements.

The above discussion is intended only to indicate the orders of magnitude of the possible sources of error and to show the feasibility of the method in principle. To check it, and to test the behaviour of the cable and the general reliability of the equipment, a series of experiments was made in which the thermistor was replaced by a 200,000 Ω resistance similar to those in the box R, and the cable was immersed in a tank whose temperature could be varied. It was found that a change of 20 °C in the temperature of the cable gave an error equivalent to approximately $0 \cdot 01$ °C.

IV. Temperature Measurements in Boreholes

The results of temperature measurements in five boreholes are recorded in Table 1 ; in all cases the holes had been drilled for some time and so were in thermal equilibrium. The depths given are corrected for stretch of the cable and, when necessary, for inclination of the hole. These results appear to represent all the useful information of this type at present available in Tasmania— measurements were also made in a number of other holes but results are not recorded here because cores were not available. The holes studied are :

Borehole No. 1.—Hole No. 5001 of the Hydro-electric Commission at the Great Lake ; height above sea-level 3378 ft ; mean annual* air temperature $5 \cdot 9$ °C. This hole is in a tholeiite (quartz-dolerite) sill with some evidence of multiple intrusion. The hole stopped at 1050 ft and did not penetrate the sill. There is no systematic jointing and no reason to believe that the hole departs from the vertical. Core recovery is complete and, as the core is being studied systematically from the petrological point of view elsewhere, we propose subsequently to make more accurate measurements of its thermal conductivity and its variation with composition. The rock may be described roughly as consisting of plagioclase and pyroxene in approximately equal amounts with a mesostasis of 10–20 per cent. Magnetite and ilmenite are occasionally present in amounts of up to 5 per cent. The grain size ranges from $0 \cdot 5$ to 3 mm.

Three independent sets of measurements, which are in excellent agreement, have been made on this hole at different times, (i) those recorded in Table 1, (ii) some measurements with maximum thermometers, and (iii) another set of

---

* In all cases the mean annual air temperatures are given merely as an instance of climate, they are not considered sufficiently reliable to use in connexion with the observations. Usually they represent averages over only a few years and are not for the actual site but for one considered to have similar climatic conditions.

readings which are not reproduced since the thermistor was not available for recalibration.

Borehole No. 2.—Hole No. 7005 of the Hydro-electric Commission on the Dee tunnel line ; height above sea-level 2412 ft ; mean annual air temperature 8·8 °C ; in tholeiite similar to No. 1.   There was a small flow of water from this hole.

Borehole No. 3.—Hole No. 7006 of the Hydro-electric Commission on the Dee tunnel line ; height above sea-level 2350 ft ; mean annual air temperature 8·8 °C ; in tholeiite similar to No. 1.

Borehole No. 4.—Hole No. 48R of the Electrolytic Zinc Company at Rosebery ; height above sea-level 651 ft ; mean annual air temperature 10·6 °C ; cased to 300 ft.

This hole and No. 5, which is close to it, are drilled through a sequence of interbedded rhyolites and rhyolitic tuffs.   In some cases phenocrysts of plagio-clase, or plagioclase in the crystal tuffs, suggest that the magma approached a dacitic composition.   Although devitrified, a few slides indicate perlitic cracking and suggest the former presence of obsidians and vitric tuffs.   All the rocks have suffered recrystallization and, in some cases, a fine groundmass consists almost entirely of interlocking grains of quartz.   A little sericite may be present in some rocks and a weak schistosity is indicated in some specimens.   Veins of quartz and of quartz-calcite are common and in some rocks vughs and veins of chlorite also occur.   Certain slides, showing silicification, probably contain as much as 90 per cent. of quartz.

Both these holes have peculiarities which make them unsatisfactory for calculation of heat flux : (i) they are not straight, the depths given in the table are corrected for the known inclination of the hole to the vertical but not for possible wandering in azimuth ; (ii) they are on a relatively steep hillside ; (iii) they are in the neighbourhood of an ore body of galena, blende, and pyrites ; (iv) large underground water movements occur in the area.

Borehole No. 5.—Hole No. 42R of the Electrolytic Zinc Company at Rosebery ; height above sea-level 866 ft ; mean annual air temperature 10·7 °C ; material similar to No. 4.   There was a small flow of water from the hole.

V. Conductivity Measurements

The thermal conductivities were measured in a divided bar apparatus similar to that described by Benfield (1939).   The brass rods were calibrated against disks of crystalline quartz cut with their plane faces parallel to the optic axis, using the value given by Griffiths and Kaye (1923) for the conductivity of quartz.   The cooling water was maintained at $26 \pm 0·02$ °C by a thermostat and the mean temperature of the specimens was about 30 °C in all cases.   The temperature gradient in each brass rod was measured by three copper-constantan thermocouples used in conjunction with a Leeds and Northrup type K2 potentio-meter.   Cores from holes 4 and 5 were 1 in. in diameter and were turned and ground down to 21 mm.   The cores of holes 1 and 2 differed in diameter from 21 mm by only 1 or 2 per cent. and were used as they stood, a correction being

TABLE 1
TEMPERATURES AND CONDUCTIVITIES IN BOREHOLES

Hole No.	Depth (ft)	Temperature (°C)	Conductivity (cal, c.g.s., °C)	Hole No.	Depth (ft)	Temperature (°C)	Conductivity (cal, c.g.s., °C)
1	56	—	0·0046	4	51	10·16	—
	101	8·63	—		54	—	0·0080
	202	10·06	0·0047		87	—	0·0081
	302	11·42	—		101	10·44	—
	343	—	0·0047		138	—	0·0097
	480	—	0·0046		150	10·77	—
	513	14·34	—		201	11·14	—
	603	15·53	—		250	11·41	—
	612	—	0·0046		300	11·73	—
	704	16·84	—		350	12·20	—
	784	—	0·0049		359	—	0·0102
	804	18·13	—		399	12·73	—
	864	—	0·0048		410	—	0·0092
	906	19·40	—		485	—	0·0077
	947	—	0·0048		497	13·70	—
	996	20·49	—		548	—	0·0099
	1033	20·91	—		588	14·67	—
					610	—	0·0057
2	20	—	0·0046		640	15·22	—
	25	9·29	—		646	—	0·0043
	50	9·49	—		687	15·78	—
	100	9·72	—				
	122	—	0·0047	5	22	—	0·0140
	151	10·17	—		47	10·60	—
	200	10·96	—		98	10·97	—
	227	—	0·0048		194	11·20	—
	252	11·58	—		242	—	0·0111
	314	—	0·0052		295	11·57	—
	316	12·38	—		393	12·43	—
					445	—	0·0098
3	25	7·72	—		493	13·33	—
	45	8·42	—		549	—	0·0101
	100	9·09	—		593	14·31	—
	151	9·78	—		645	—	0·0093
	202	10·46	—		691	15·16	—
	252	11·13	—		736	—	0·0118
					742	—	0·0050
					787	16·04	—
					844	—	0·0058
					851	—	0·0060
					877	16·98	—
					898	17·17	—

made for the error in diameter. This correction has been studied in detail both theoretically and experimentally and will be described elsewhere. The cores from hole No. 3 were only ¾ in. in diameter and so could not be measured in this apparatus. Four sections, 1, 2, 4, and 6 mm in thickness, were cut for each experiment, the 1 mm section being used to make a slide for petrological examina-

tion and the others being ground flat to within $0 \cdot 0005$ in. In most cases a film of glycerol was used at the contact surfaces.

Values of the conductivities are given in Table 1, the units being c.g.s., calorie, and °C. It is believed that these are accurate to within 4 per cent. Since we find a variation of this order between samples a few inches apart in the core, there is no point in pushing the accuracy further, though an attempt will be made with an improved apparatus to correlate the variations of conductivity with petrological properties for the case of hole No. 1.

VI. THE HEAT FLUX

In all cases this has been estimated by plotting temperature against thermal resistance from the surface. Results for the individual holes are as follows.

Borehole No. 1.—The value of the heat flux is $2 \cdot 04 \times 10^{-6}$ cal cm^{-2} sec^{-1}.

Borehole No. 2.—The anomalous behaviour at the lower depths is presumably caused by the flow of water mentioned. Using only the results for depths greater than 100 ft gives a value of $2 \cdot 06 \times 10^{-6}$ for the heat flux.

Borehole No. 3.—In this case, as mentioned above, conductivities were not measured but, since the hole is close to No. 2, a reasonable value of the heat flux may be obtained by using the observed thermal gradient in conjunction with the harmonic mean of the conductivities measured in hole No. 2. This gives a value of $2 \cdot 07 \times 10^{-6}$.

Borehole No. 4.—The temperature-depth curves both of this hole and No. 5 suggest that they are " two-layer " holes but in fact the rocks are of the same type throughout, though there are large variations both of composition and thermal conductivity. The heat flux in this case is $2 \cdot 47 \times 10^{-6}$.

Borehole No. 5.—The first 300 ft have been disregarded because of the water flow mentioned above. The lower depths give a heat flux of $2 \cdot 54 \times 10^{-6}$.

The general conclusion drawn from these results is that the heat flux in Tasmania has the relatively high value of 2×10^{-6} cal cm^{-2} sec^{-1}. This is based on hole No. 1 alone which has been very carefully investigated. It is unfortunate that no other hole which could be regarded as satisfactory from the point of view of thermal measurements was available and, for this reason, we have included the other four holes as corroborative evidence. Nos. 2 and 3 confirm this value satisfactorily and Nos. 4 and 5, being in the neighbourhood of a sulphide ore body, would be expected to give somewhat higher values.

These values do not allow for the influence of the recent ice age on the present heat flux. It is usual to estimate this roughly by assuming that the surface assumed its present mean annual temperature V immediately on the conclusion of the ice age : if t is the time since this conclusion, and k, ρ, and c are the thermal conductivity, density, and specific heat of the surface material, the heat flux at depth will be greater than that at the surface by an amount

$$V(k\rho c/\pi t)^{\frac{1}{2}},$$

which, with the values $V = 5 \cdot 9$ °C for hole No. 1, $\rho c = 0 \cdot 5$, $k = 0 \cdot 0050$, gives an addition of $0 \cdot 42 \times 10^{-6}$ to the heat flux if t is given the reasonable value of 5000 years.

VII. Acknowledgments

The authors are greatly indebted to the Hydro-electric Commission of Tasmania and the Electrolytic Zinc Company for putting many facilities at our disposal and for the ready cooperation of members of their staffs. Thanks are also due to Professor S. W. Carey of the University of Tasmania for advice on the selection of boreholes, to Dr. Germaine Joplin of the Australian National University for the petrological descriptions of the specimens, and to Professor J. C. Jaeger of the Australian National University for suggesting the construction and applications of the instrument.

VIII. References

BENFIELD, A. E. (1939).—*Proc. Roy. Soc.* A **173** : 428.

DEETER, E. L. (1948).—*Electronics* **21** : 180.

DOUCET, Y. (1951).—*J. Phys. Radium* **12** : 823.

GRIFFITHS, E., and KAYE, G. W. C. (1923).—*Proc. Roy. Soc.* A **104** : 71.

LYNCH, L. E. V., and ROBERTSON, D. S. (1946).—*A.W.A. Tech. Rev.* **1** : 8.

MISENER, A. D., THOMPSON, L. G. D., and UFFEN, R. J. (1951).—*Trans. Amer. Geophys. Un.* **32** : 729.

MONROE, A. G., and PENNY, D. J. (1952).—*J. Sci. Instrum.* **29** : 163.

STANDARD TELEPHONES AND CABLES LTD. (1952).—Thermistor Application Report, January 1952.

VALLEY, G. E., and WALLMAN, H. (1948).—" Vacuum Tube Amplifiers." p. 384. M.I.T. Radiation Laboratory Series, Vol. 18. (McGraw-Hill : New York.)

I

SHORT COMMUNICATIONS

THE INFLUENCE OF METEORITIC DUST ON RAINFALL*

By E. G. Bowen†

A great deal of attention has been given in the meteorological literature to periodicities in the weather and to singularities in temperature or pressure, that is, occasions on which these quantities tend to have an abnormally high or an abnormally low value. The analysis of daily rainfall figures has not received as much attention, however, and it is the purpose of this note to point out some unexpected features in the rainfall occurring in different parts of the world and to advance a possible explanation for the phenomenon.

The weekly or monthly mean rainfall figures for any one station over a number of years seldom show departures from the mean which are greater than would be expected from statistical fluctuations. If the daily rainfall figures are examined, however, it is apparent that in some localities there is a marked tendency for heavy falls of rain to occur on certain days rather than on others, and for this pattern to be repeated year after year.
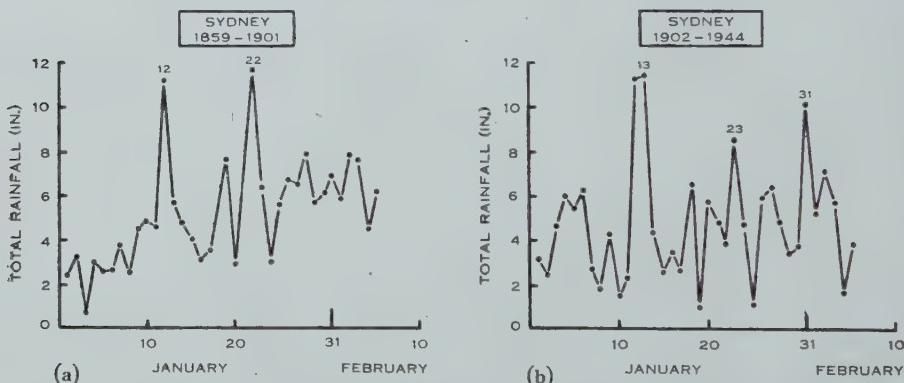
Fig. 1.—Daily rainfall of Sydney for January. (a) 1859–1901 and (b) 1902–1944.

For purposes of illustration the total daily rainfall of Sydney for January and the first five days of February for the period from 1859 to 1901 is given in Figure 1 (a). It shows a general increase throughout the month, conforming to the seasonal pattern, and two exceptional peaks occurring respectively on January 12 and January 22. The magnitude of each peak is approximately twice the mean and their departure from the mean about four times the standard

deviation of the remainder. The corresponding curve for the period from 1902 to 1944 is given in Figure 1 (b) and it is seen that two similar peaks occur on January 12–13 and January 23, while another appears on January 31. It will be shown later than this third peak is a significant one.

The near coincidence in time of the first two peaks was unexpected enough to stimulate an examination of the rainfall figures of other stations in corresponding latitudes to see whether they too showed any unusual characteristics. The surprising discovery was made that many stations over a wide area tended to show peaks of similar magnitude on nearly the same days. Figures for periods of approximately 50 years are available for seven stations in the southern hemisphere extending over 180° in longitude from South Africa to New Zealand. These are listed in Table 1, together with the dates on which the rainfall exceeds the mean by 50 per cent.
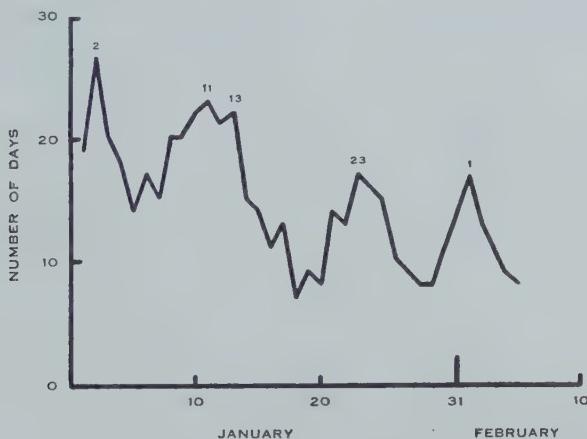
Fig. 2.—Number of occasions in January on which heavy falls of rain were recorded in the British Isles from 1919 to 1949.

Figures for a comparable area in the northern hemisphere are, unfortunately, not available at the time of writing, but records are available for a few individual stations in the British Isles. They show that the rainfall there has a smaller variability than in the stations just discussed and they do not show peaks which depart significantly from the normal fluctuations. Examination of the daily records in localities which do show the peaks of rainfall indicates that they are due to a comparatively few days of exceptionally heavy rain and are not due to a greater frequency of rain on those days. The British records have therefore been re-examined, taking into account only exceptionally heavy falls. These are recorded in the publication " British Rainfall " and Figure 2 gives the number of days on which heavy* rain fell anywhere in the British Isles during the period from 1919 to 1949. This curve shows a maximum on January 2 which does not correspond to those already discussed and three other maxima

* Heavy rain is defined in the records as more than $2\frac{1}{2}$ in. in a day, or more than $7\frac{1}{2}$ per cent. of the annual rainfall in areas where this is less than $33 \cdot 3$ in.

II

respectively on January 11, January 23, and February 1, corresponding closely to those in the Sydney rainfall.

These peaks have been included with those of southern hemisphere stations in Table 1. It is seen that in all stations there is a remarkable grouping of the peaks, the actual distribution over the month being given in Figure 3. With the single exception of January 2 already noted, the peaks are grouped around January 13, 23, and 31 with a spread of ± 2 days about these dates. Although these examples are confined to January, similar results are obtained in other months of the year.

TABLE 1

DATES OF RAINFALL PEAKS

Place	Period	Dates on which Peaks of Rain Occur			
Durban	1900–53		Jan. 12 & 16	Jan. 22	Jan. 30
Perth	1907–52		15	20 & 24	31
Alice Springs ..	1900–52		14	21	28 & 31
Sydney	1900–49		12 & 13	23	31
Brisbane	1900–49		13	24	Feb. 1
Auckland	1900–53		13	21	Jan. 29 & Feb. 1
Christchurch ..	1905–53		14	24	—
Great Britain ..	1919–49	Jan. 2	11 & 13	23	Feb. 1
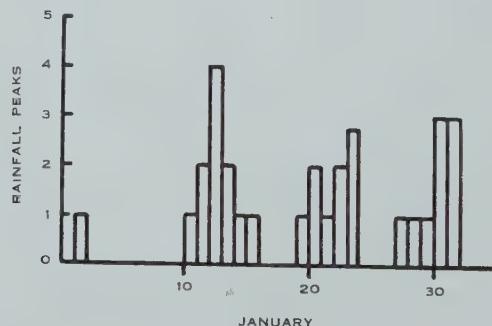
Fig. 3.—Distribution of rainfall peaks in January
for Durban, Perth, Alice Springs, Brisbane, Sydney,
Auckland, Christchurch, and the British Isles.

The question arises whether this phenomenon can be accounted for by climatological factors alone or whether some other influence is at work. If it were due to climatological effects one would expect it to be propagated with a velocity similar to those of weather systems. No such tendency is found and there is no evidence for progressive displacement in the timing of the peaks with geographical location. Furthermore, the particular years on which heavy falls of rain contributed to the peak on January 14, at Alice Springs for example, are quite different from those which gave a peak on the same date at Christchurch. Finally, it is difficult to conceive of any climatological effect which would give

heavy rainfall in the southern hemisphere on approximately the same dates as in the northern hemisphere. Clearly, the explanation of the effect is likely to be found in a phenomenon which can operate simultaneously over a large part of the world ; this suggests that it is of extraterrestrial origin. Of the great number of extraterrestrial phenomena which could be responsible, the only one which meets the requirement of repeating year after year on the same dates is that of meteor showers.

The average number of meteors which can be seen by eye in any part of the sky is 10 or 12 per hour. These are due to meteor particles which are distributed more or less at random in the solar system. Occasionally, however, the Earth passes through vast meteor streams which follow elliptical orbits about the Sun and the number of visible meteors can then increase to upwards of 50 or 60 an hour. Ten night-time streams of this kind are known to observers in the northern hemisphere and they occur regularly year after year at fixed dates and times. Other, daylight streams have recently been discovered by radar methods and a complete description can be found in the literature (see Lovell and Clegg 1952 ; Porter 1952).

The dates of meteor streams nearest in time to the rainfall peaks already discussed are :

Geminids	December 13–14
Ursids	December 22
Quadrantids	January 3

That is, the rainfall peaks tend to follow the meteor showers after an interval of about 30 days.

Meteor streams occur predominantly at two periods of the year, during May, June, and July and again during October, November, and December. Those in May, June, and July are closely grouped in time and are difficult to distinguish from one another. Those in October, November, and December, however, are separated by longer intervals. If the Sydney rainfall for this period is examined it is found that distinct peaks of rainfall follow each meteor shower with an average delay of about 29 days. In Figure 4 is given the curve of the daily rainfall during November, December, and January from 1900 to 1949. Immediately above are the dates of each meteor shower displaced 29 days in time. The correspondence with the rainfall peaks is obviously a close one.

It can be concluded therefore that at certain stations and at certain times of the year there is a high probability of a peak of rainfall appearing some 29 or 30 days after the Earth enters a major meteor stream.

It remains to explore the physical process by which meteor showers might influence world rainfall in this way. One possibility is that the meteoritic dust accompanying meteors provides rain-forming nuclei when it falls into cloud systems in the lower atmosphere. It is well known that the formation of rain in clouds is dependent on the existence of certain types of nuclei. On occasion, these are not available from terrestrial sources and clouds build up to great heights without giving rain.

The dust in interplanetary space is known to consist of particles with a wide range of sizes. These are swept up by the Earth in its orbit, and Whipple (1950) has shown that those greater than 4 µ in diameter have sufficient energy to burn up on entering the atmosphere and become visible as meteors. Those smaller than 4 µ are not consumed and fall slowly to the ground. The total mass of material falling on the Earth in sizes large enough to give visible meteors is estimated to be about a ton a day. The amount of accompanying material in smaller sizes appears to be very much greater and van de Hulst (1947) has shown, from considerations of the way in which they scatter light from the
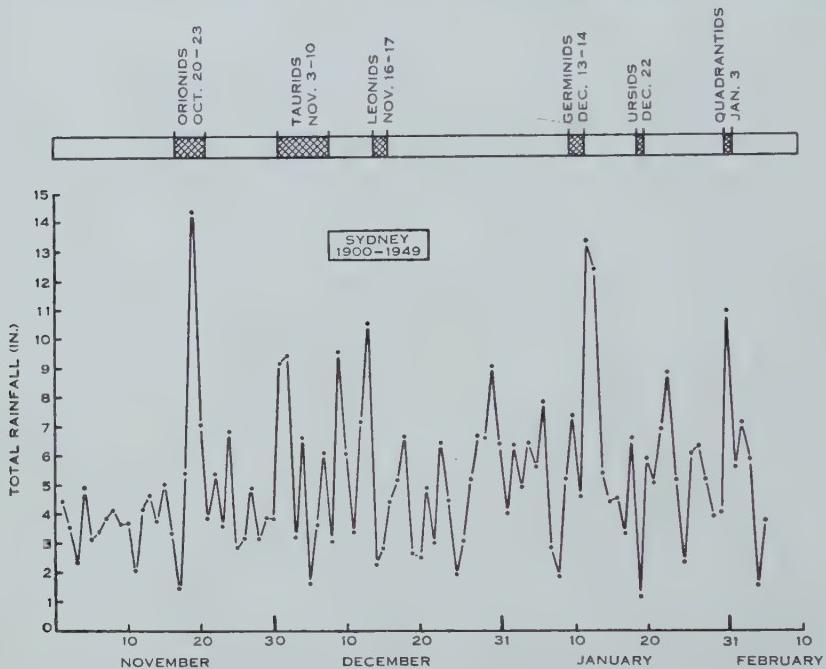
Fig. 4.—Daily rainfall of Sydney for November, December, and January for the period 1900–1949, together with dates of the main meteor streams displaced 29 days in time.

Sun, that the concentration of these particles in free space is approximately $10^{-6}/m^3$, and the total mass swept up by the Earth is on the average about 10,000 tons a day.

The particles have velocities ranging from 10 to 30 km/sec in relation to the Earth and on entering the atmosphere they are decelerated suddenly at about 100 km, a height which is relatively independent both of their initial velocity and direction of arrival. Their velocity falls abruptly from tens of kilometres a second to less than one centimetre a second. A sudden increase of particle density of about a million times therefore takes place at this level; that is, their concentration within the atmosphere will approach $1/m^3$.

A calculation of the time of fall of particles of 1–4 µ diameter from the 100 km level shows that they would take from 30 to 50 days to descend to the

40,000 or 50,000 ft level.   At this height they could meet some of the larger cloud structures of the lower atmosphere and might thus influence the incidence of rainfall.

It is seen therefore that meteoritic dust exists in adequate quantities to affect the rainfall of the lower atmosphere and its time of fall is of the right order to account for the observed interval between meteor showers and peaks of rainfall.
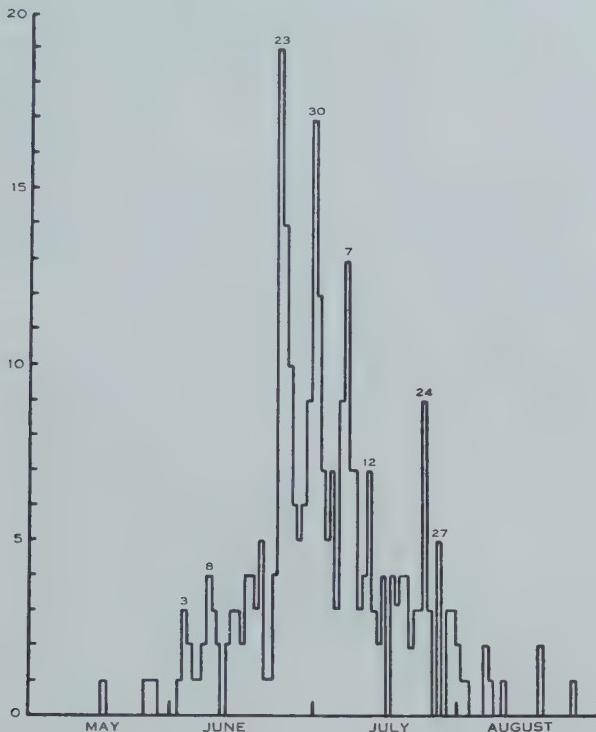
Fig. 5.—Number of occasions on which noctilucent clouds have been observed for the period 1885–1933 (Vestine 1934).

The Relation between Meteoritic Dust and Noctilucent Clouds

If, in fact, meteoritic dust can influence rainfall in this way it might produce other effects in the atmosphere which can be recognized.   Among the effects already reported is a reduction in atmospheric transparency about the time of the Perseids described by Zacharov (1952).   The ionization produced by the heavier particles in the ionosphere is, of course, well known and will not be enlarged upon.

The sudden stoppage of the smaller particles and the corresponding increase in their concentration in the 80–100 km region leads one to connect them with noctilucent clouds, which are occasionally seen at these heights.   The origin of these clouds has never been completely determined but Vestine (1934) examined the possibility of their being due to dust from volcanoes, the debris

of comets, or meteoritic dust. He concluded that there was no evidence in favour of the first two and that they were probably due to the third.

This probability becomes almost a certainty if a comparison is made between the incidence of noctilucent clouds and the time of occurrence of meteor streams. Vestine gives a curve, which is reproduced in Figure 5, of the total number of occasions on which noctilucent clouds have been seen in the period from 1885 to 1933. These are concentrated around the summer solstice, when seeing conditions are good in the northern hemisphere. This curve has a number of well-defined peaks which occur on the dates given in the first column of Table 2. In the second column are given the dates of the principal meteor streams which occur during the same period. It is seen at once that the noctilucent clouds

TABLE 2

COMPARISON BETWEEN DATES OF APPEARANCE OF
NOCTILUCENT CLOUDS AND METEOR SHOWERS

Dates on which Noctilucent Clouds Observed (Vestine)	Date of Meteor Shower (Lovell)	Name of Meteor Shower
June 3	June 3	ζ-Perseids
8	8	Arietids
23	25	54-Perseids
30	July 2	β-Taurids
July 7	—	—
12	12	ν-Geminids
24	25	θ-Aurigids
27	28	δ-Aquarids

tend to appear either at precisely the same date or within a few days of the meteor streams. It is noteworthy that, with the exception of the δ-Aquarids on July 28, the whole of these are daylight streams which were not discovered until 1948 and were unknown when Vestine examined the data.

It can be concluded with a fair degree of certainty, therefore, that noctilucent clouds have their origin in meteoritic dust. Whether the material which becomes visible is the dust itself or whether it acts as a nucleus for the formation of ice crystals or the condensation of water vapour remains to be decided. Noctilucent clouds have been seen at line-of-sight distances up to 600 km and the same cloud has been seen simultaneously from Canada and Siberia. Clearly a great number of particles are involved and when they fall to the lower atmosphere they might well influence the rainfall in the manner which has been discussed.

Meteorological Implications of the Phenomenon

In conclusion, some brief remarks are made about the meteorological implications of the phenomenon discussed in this paper.

In the first place, the art of forecasting might be influenced in one important respect. As pointed out earlier, while the amount of rain on a certain date

might be considerably in excess of normal, the probability of rain occurring on that day is no higher than on the days immediately preceding or following it. In other words, the probability that clouds suitable for the formation of rain will build up is determined by climatological factors and, averaged over a long period of years, would not show great variations from one day to the next. The amount of rain, however, is influenced by the supply of rain-forming nuclei in the atmosphere. In the absence of nuclei from terrestrial sources, knowledge of an impending fall of meteoritic dust might be important in estimating whether extremes of rain are likely to be obtained in any given climatological situation.

Finally, the results might have an important bearing on the problem of artificial rainmaking. If the present results are substantiated it means that the presence of relatively small concentrations of particles falling from the upper atmosphere can result in the rainfall of particular days being about double that of days preceding them. It suggests that, at least in certain parts of the globe, the potential increase in rainfall which could be obtained by artificial methods might exceed the figure of 10 or 15 per cent. which has previously been estimated.

References

VAN DE HULST, H. C. (1947).—*Astrophys. J.* **105** : 471-88.
LOVELL, B., and CLEGG, J. A. (1952).—" Radio Astronomy." (Chapman and Hall Ltd.: London.)
PORTER, J. G. (1952).—" Comets and Meteor Streams." (Chapman and Hall Ltd. : London.)
VESTINE, E. H. (1934).—*J. R. Astr. Soc. Can.* **28** : 249-72.
WHIPPLE, F. L. (1950).—*Proc. Nat. Acad. Sci.* **36** : 687-95.
ZACHAROV, I. (1952).—*Bull. Central Astr. Inst. Czechoslovakia* **3** : 82-5.

NOTE ON THE FACTORIAL MOMENTS OF STANDARD DISTRIBUTIONS*

By R. B. Potts†

The factorial moments of the standard discontinuous distributions can be readily derived by means of a generating function, and it is well known that the formulae for these moments are much simpler than those for the more usual power moments. It is the purpose of the present note to draw attention to an interesting alternative method of deriving these factorial moments.

If $\varphi(x)$ is the distribution function of a discrete variate x which is capable of only integral values, then the rth factorial moment about the origin is defined to be

$$T(r) = \sum_{j=0}^{\infty} \frac{(r+j)!}{j!} \varphi(r+j) \quad \dots\dots\dots\dots\dots (1)$$

(the summation may be finite). The formula inverse to (1) is

$$\varphi(x) = \sum_{j=0}^{\infty} \frac{(-1)^j}{x!j!} T(x+j), \quad \dots\dots\dots\dots\dots (2)$$

which expresses the distribution function $\varphi(x)$ in terms of the xth and higher factorial moments. It is this relation which will be used in deriving the factorial moments of the standard discontinuous distributions ; there is, incidentally, no relation as simple as (2) which expresses the distribution function in terms of the power moments.

Binomial Distribution

For the binomial distribution,

$$\varphi(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad \dots\dots\dots\dots\dots (3)$$

which gives, by expanding $(1-p)^{n-x}$,

$$\varphi(x) = \sum_{j=0}^{n-x} \frac{(-1)^j}{x!j!} \frac{n!}{(n-x-j)!} p^{x+j}. \quad \dots\dots\dots\dots (4)$$

By comparison with (2), the factorial moments of the binomial distribution are given by

$$T(r) = \frac{n!}{(n-r)!} p^r. \qquad\qquad (5)$$

Poisson Distribution

For the Poisson distribution,

$$\varphi(x) = \frac{m^x e^{-m}}{x!} = \sum_{j=0}^{\infty} \frac{(-1)^j}{x!j!} m^{x+j}, \qquad\qquad (6)$$

by expanding the exponential. Hence, from (2), the factorial moments are

$$T(r) = m^r. \qquad\qquad (7)$$

Polya Distribution

For this distribution,

$$\varphi(x) = (1+b)(1+2b)\ldots\{1+(x-1)b\}a^x(1+ab)^{-x-b^{-1}}/x!, \qquad (8)$$

from which one obtains

$$\varphi(x) = \sum_{j=0}^{\infty} \frac{(-1)}{x!j!} (1+b)(1+2b)\ldots\{1+(x+j-1)b\}a^{x+j}. \qquad (9)$$

Hence, from (2), the factorial moments of the Polya distribution are

$$T(r) = (1+b)(1+2b)\ldots\{1+(r-1)b\}a^r. \qquad\qquad (10)$$

Hypergeometric Distribution

The distribution function is, in the usual notation,

$$\varphi(x) = \binom{Np}{x} \binom{N-Np}{n-x} \bigg/ \binom{N}{n}. \qquad\qquad (11)$$

From the identity

$$\binom{N-Np}{n-x} = \sum_{j=0}^{n-x} (-1)^j \binom{Np-x}{j} \binom{N-x-j}{n-x-j}, \qquad (12)$$

it follows that

$$\varphi(x) = \sum_{j=0}^{n-x} \frac{(-1)^j}{x!j!} \frac{(Np)!}{(Np-x-j)!} \frac{n!}{(n-x-j)!} \frac{(N-x-j)!}{N!}, \quad (13)$$

and therefore, from (2),

$$T(r) = \frac{(Np)!}{(Np-r)!} \frac{n!}{(n-r)!} \frac{(N-r)!}{N!}. \qquad\qquad (14)$$

Recent work on number distributions occurring in cosmic ray cascade theory has revealed the especial importance of the factorial moments, and such relations as (2), for discontinuous distributions which are more complex than the simple standard ones considered above.

ANOMALOUS ABSORPTION OF COSMIC RAYS IN LEAD*

By W. L. KENNEDY†

Anomalies in the lead absorption curve of the vertical component of cosmic radiation have been reported by several authors using counter telescopes (Aiya 1944 ; George and Appapillai 1945 ; Swann and Morris 1947 ; Kellermann and Westerman 1949 ; Fenyves and Haiman 1950 ; Mazzolli de Mathov 1951 ; Abd El-Wahab Khalil 1952). However, in a recent detailed series of experiments, results of which were published as the experiment to be described here was nearing completion, Heyland and Duncanson (1953) found no evidence of any anomaly in the absorption curve.

In the present experiment, the absorption of the vertical component of cosmic radiation in lead was measured at sea-level, geomagnetic latitude 43·5 °S., using narrow-angle triple-coincidence counter telescopes. Details of the
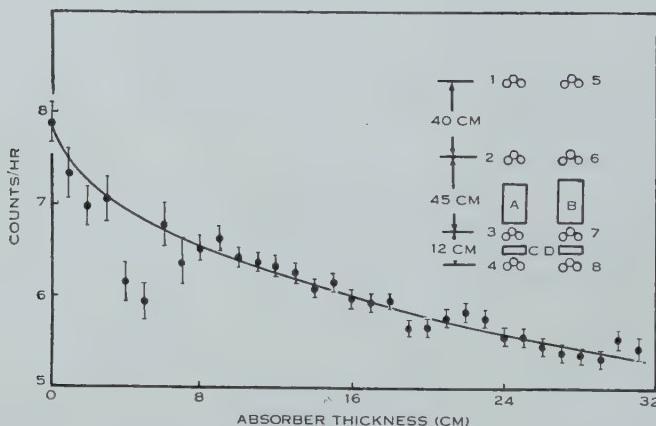
Fig. 1.—Absorption curve of cosmic rays in lead, showing counter arrangement inset.

arrangement of the counters are inset in Figure 1. Four telescopes, consisting of trays 123, 124, 567, and 568, were in operation simultaneously, and results for four adjacent absorber thicknesses were obtained by having $C=D=1$ cm lead, and $B=A+2$ cm. The effective area of each counter tray was 10·6 by 5·7 cm, and the angular aperture of each telescope was thus approximately 13 by 7°.

Pressure and temperature coefficients were determined to be $-3 \cdot 0$ per cent./cm Hg and $-0 \cdot 20$ per cent./°C for absorber thickness of 10 cm, and these coefficients were applied as corrections to observations at other absorber thicknesses.

Before combining the results of the various telescopes, a check on relative efficiency was made. As the observed counting rates at the same absorber thickness were very closely in the ratio of the angular apertures, it was concluded that efficiencies were equal, and only a geometrical aperture correction was required. In this connexion, it followed that the shower counting rate was small and also, in view of the results obtained by Swan (1951), no corrections for showers were considered necessary. The accidental triple-coincidence counting-rate was less than $0 \cdot 1$ counts/day, and therefore negligible.

The final results of the absorption measurements are presented in Figure 1. Each point is based on approximately 2000 counts, except for those at less than 8 cm absorber thickness, for these latter were considered to be of lesser importance in the present investigation.

Analysis of the absorption curve of Figure 1 indicates that there are a number of small deviations from the smooth curve occurring in the region from 8 to 31 cm. It appears unlikely that these deviations result entirely from neglect of a shower correction or inaccuracies in applying barometric and temperature corrections ; these should of themselves introduce no maxima or minima in the absorption curve.

As none of the deviations is statistically significant, it is concluded that the present results are in agreement with the assumption of a smooth absorption curve.

References

ABD EL-WAHAB KHALIL, M. (1952).—*Nuovo Cim.* **9** : 1248.

AIYA, S. V. C. (1944).—*Nature* **153** : 375.

FENYVES, E., and HAIMAN, O. (1950).—*Nature* **165** : 244.

GEORGE, E. P., and APPAPILLAI, V. (1945).—*Nature* **155** : 726.

HEYLAND, G. R., and DUNCANSON, W. E. (1953).—*Proc. Phys. Soc. Lond.* A **66** : 33.

KELLERMANN, E. W., and WESTERMAN, K. (1949).—*Proc. Phys. Soc. Lond.* A **62** : 356.

MAZZOLLI DE MATHOV, E. (1951).—*Nature* **167** : 192.

SWAN, J. B. (1951).—*Nature* **168** : 73.

SWANN, W. F. G., and MORRIS, P. F. (1947).—*Phys. Rev.* **72** : 1262.

^{8}Li(β)^{8}Be*(2α) HAMMER TRACKS IN PHOTOGRAPHIC EMULSIONS†

By A. R. W. Wilson‡

The α-pair energy spectrum of 94 ^{8}Li(β)^{8}Be*(2α) hammer tracks observed in photographic emulsion and reported in the literature has been shown by Pickup and Voyvodic (1950) to agree with the results of Bonner *et al.* (1948), and Bennett, Bonner, Mandeville, and Watt (unpublished data 1948, see Bonner *et al.* 1948), who used cloud chambers to measure the pair energy spectrum up to
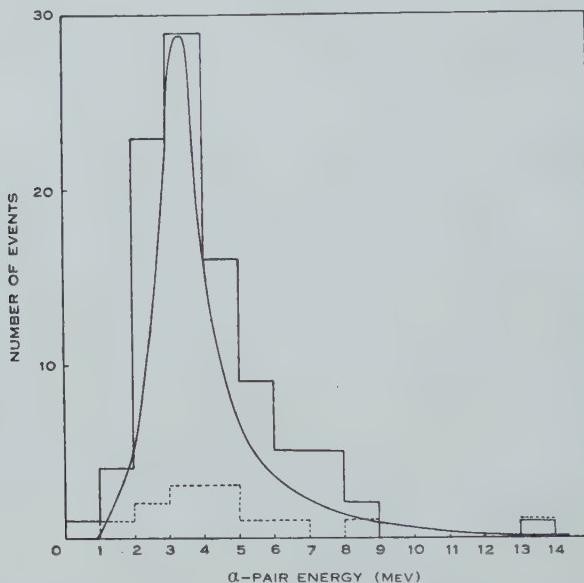
Fig. 1.—Energy spectrum of pairs from ^{8}Li(β)^{8}Be*(2α) hammer tracks.
– – – – The 13 events found in this laboratory.
———— The 107 events obtained by summing the results of Franzinetti and Payne (1948), Pickup and Voyvodic (1950), and Titterton (1950), together with the present results.
The smooth curve is drawn through the experimental results of Bonner *et al.* (1948) over the 0–10 MeV range and of Rumbaugh, Roberts, and Hafstad (1938) above the latter energy. It is normalized to agree with the solid line histogram at its maximum.

about 10 MeV. Smith and Chang (1938), using counters, observed α-particles up to an energy of $E_\alpha = 7 \cdot 75$ MeV, and Rumbaugh, Roberts, and Hafstad (1938) α-pairs up to an energy of $15 \cdot 8 \pm 0 \cdot 5$ MeV; however, the number of disintegrations giving an α-pair energy greater than 9 MeV was $\leqslant 1$ per cent. From an

analysis of the β-spectrum, Hornyak and Lauritsen (1950) have suggested that about 10 per cent. of the ^{8}Li transitions should go to ^{8}Be states at ~10 MeV and ~13 MeV. The small fraction of α-pairs with energies near these levels observed experimentally may be due to the states undergoing γ-ray transitions. Hammer tracks with α-pair energies greater than 9 MeV have not previously been reported as observed in photographic emulsions.

Thirteen hammer tracks have been found in 4000 stars in G5 plates exposed at approximately 80,000 ft and scanned by the cosmic ray group in this laboratory. The broken lines (Fig. 1) show the energy spectrum of the α-pairs ; it includes one pair with an energy of 13 MeV from an ^{8}Li fragment emitted from a light nucleus.

The 13 α-pairs have been combined with the 94 previously reported to give the full-line histogram of Figure 1. This is consistent with the energy spectrum obtained by drawing a smooth curve through the cloud chamber results of

TABLE 1

ANALYSIS OF 13 ^{8}Li(β)^{8}Be*(2α) HAMMER TRACKS

α-Pair Energy (MeV)	Identification of β-Particles	β-Energy (MeV)	Angle between β and α	Angle between ^{8}Li and α	^{8}Li Energy (MeV)	Number of Evaporation Tracks in Star
1·8	Certain	7	88°	88°	23	3
6·1	,,	~1	62°	88°	20	16
3·3	,,	~2	17°	33°	18	8
5·2	Doubtful	—	—	82°	11	11
13·1	,,	—	—	76°	3	4
4·1	Certain	10	79°	62°	6	4
2·7	Not observed	—	—	75°	23	3
4·3	Doubtful	—	—	64°	20	8
3·9	Certain	~2	55°	71°	9	4
8·6	,,	>1	34°	47°	44	11
2·6	,,	8	76°	36°	24	14
3·3	,,	10	46°	88°	13	19
4·1	Not observed	—	—	47°	14	13

Bonner et al. (1948) up to 10 MeV and the counter results of Rumbaugh, Roberts, and Hafstad (1938) above that energy and normalizing the curve to agree with the block spectrum at its maximum.

In eight of the 13 events, the track of the ^{8}Li decay electron was clearly observable and, in a further three events, what was probably a low energy electron track was observed coming from the vicinity of the ^{8}Be break-up point. The absence of electron tracks in the other two cases could reasonably be attributed in one case to the proximity of the break-up point to the emulsion surface and in the other to β-emission in a direction unfavourable for observation. Table 1 sets out the data from the analysis of the 13 hammer tracks.

The β-α angular correlation of the hammer tracks coming from the 3 MeV state of ^{8}Be could be useful in deciding the spin and parity of ^{8}Li. Gardner

(1951) has shown that if the ^{8}Li has spin and parity $0+$ the β-α angular distribution should show a maximum near 45°, particularly for high energy β-particles. If the ^{8}Li spin and parity is $3-$, the distribution, whilst still anisotropic, should not have a maximum between 0 and 90°. The statistics of the present events are insufficient to distinguish between these cases but could be added to the results of other laboratories.

References

BONNER, T. W., EVANS, J. E., MALICH, C. W., and RISSER, J. R. (1948).—*Phys. Rev.* **73** : 885.

FRANZINETTI, C., and PAYNE, R. M. (1948).—*Nature* **161** : 735.

GARDNER, J. W. (1951).—*Phys. Rev.* **82** : 283.

HORNYAK, W. F., and LAURITSEN, T. (1950).—*Phys. Rev.* **77** : 160.

PICKUP, E., and VOYVODIC, L. (1950).—*Canad. J. Res.* A **28** : 616.

RUMBAUGH, L. H., ROBERTS, R. B., and HAFSTAD, L. R. (1938).—*Phys. Rev.* **54** : 657.

SMITH, C. L., and CHANG, W. Y. (1938).—*Proc. Roy. Soc.* A **166** : 415.

TITTERTON, E. W. (1950).—*Nature* **165** : 721.

INDEX